Sponsored by the Association for Education Finance and Policy (AEFP), the second edition of this groundbreaking handbook assembles in one place the existing research-based knowledge in education finance and policy, with particular attention to elementary and secondary education. Chapters from the first edition have been fully updated and revised to reflect current developments, new policies, and recent research. With new chapters on teacher evaluation, alternatives to traditional public schooling, and cost-benefit analysis, this volume provides a readily available current resource for anyone involved in education finance and policy.

The *Handbook of Research in Education Finance and Policy* traces the evolution of the field from its initial focus on school inputs and revenue sources used to finance these inputs, to a focus on educational outcomes and the larger policies used to achieve them. Chapters show how decision making in school finance inevitably interacts with decisions about governance, accountability, equity, privatization, and other areas of education policy. Because a full understanding of important contemporary issues requires inputs from a variety of perspectives, the *Handbook* draws on contributors from a number of disciplines. Although many of the chapters cover complex, state-of-the-art empirical research, the authors explain key concepts in language that nonspecialists can understand. This comprehensive, balanced, and accessible resource provides a wealth of factual information, data, and wisdom to help educators improve the quality of education in the United States.

**Helen F. Ladd** is the Susan B. King Professor of Public Policy Studies and Professor of Economics at the Sanford School of Public Policy at Duke University.

**Margaret E. Goertz** is Senior Researcher at the Consortium for Policy Research in Education and Professor Emerita of Education Policy in the Graduate School of Education at the University of Pennsylvania.
Contents

Preface

Acknowledgments

Section I PERSPECTIVES ON EDUCATION FINANCE AND POLICY
MARGARET E. GOERTZ AND HELEN F. LADD, SECTION EDITORS

Chapter 1 History and Scholarship Regarding U.S. Education Finance and Policy
MATTHEW G. SPRINGER, ERIC A. HOUCK, AND JAMES W. GUTHRIE

Chapter 2 The Role of Economics in Education Policy Research
DOMINIC J. BREWER, GUILBERT C. HENTSCHKE, AND ERIC R. EIDE, WITH TENICE HARDAWAY AND TIEN LE

Chapter 3 The Past, Present, and Possible Futures of Educational Finance Reform Litigation
WILLIAM S. KOSKI AND JESSE HAHNEL

Chapter 4 The Continually Evolving Political Context of Education Finance
JAMES W. GUTHRIE AND KENNETH K. WONG

Chapter 5 Educational Goals: A Public Perspective
REBECCA JACOBSEN AND RICHARD ROTHSTEIN

Chapter 6 Quantitative Research Methods in Education Finance and Policy
PATRICK J. McEWAN

Chapter 7 International Large-Scale Assessments: Uses and Implications
STEPHEN P. HEYNEMAN AND BOMMI LEE

Section II MAKING MONEY MATTER
JENNIFER KING RICE, SECTION EDITOR

Chapter 8 Toward an Understanding of Productivity in Education
JENNIFER KING RICE AND AMY ELLEN SCHWARTZ

Chapter 9 Cost-Benefit Analysis and Cost-Effectiveness Analysis
CLIVE R. BELFIELD

Chapter 10 Teachers Clearly Matter, But Finding Effective Teacher Policies Has Proven Challenging
DAN GOLDHABER

Chapter 11 Teacher Evaluation for Accountability and Development
MORGAEN L. DONALDSON AND JOHN P. PAPAY

Chapter 12 School Accountability and Student Achievement
DAVID N. FIGLIO AND HELEN F. LADD

Chapter 13 School Competition and Student Outcomes
BRIAN GILL AND KEVIN BOOKER
Section III PROMOTING EQUITY AND ADEQUACY
LEANNA STIEFEL, SECTION EDITOR

Chapter 14 Conceptions of Equity and Adequacy in School Finance
BRUCE D. BAKER AND PRESTON C. GREEN

Chapter 15 Measuring Equity and Adequacy in School Finance
THOMAS A. DOWNES AND LEANNA STIEFEL

Chapter 16 Measurement of Cost Differentials
WILLIAM D. DUNCOMBE, PHUONG NGUYEN-HOANG, AND JOHN YINGER

Chapter 17 Intergovernmental Aid Formulas and Case Studies
LAWRENCE O. PICUS, MARGARET E. GOERTZ, AND ALLAN R. ODDEN

Chapter 18 Education Equity in an International Context
EDWARD B. FISKE AND HELEN F. LADD

Section IV CHANGING PATTERNS OF GOVERNANCE AND FINANCE
ANDREW RESCHOVSKY, SECTION EDITOR

Chapter 19 The Changing Federal Role in Education Finance and Governance
NORA E. GORDON

Chapter 20 The Role of Nongovernmental Organizations in Financing Public Schools
JANET S. HANSEN, MICHELLE HALL, DOMINIC J. BREWER, AND JANE HANNAWAY

Chapter 21 Equity, Adequacy, and the Evolving State Role in Education Finance
SEAN P. CORCORAN AND WILLIAM N. EVANS

Chapter 22 Local Funding of Schools: The Property Tax and Its Alternatives
THERESE J. MCGUIRE, LESLIE E. PAPKE, AND ANDREW RESCHOVSKY

Chapter 23 Tax and Expenditure Limits, School Finance, and School Quality
THOMAS A. DOWNES AND DAVID N. FIGLIO

Section V EDUCATIONAL MARKETS AND DECENTRALIZATION
HENRY M. LEVIN AND RON ZIMMER, SECTION EDITORS

Chapter 24 Issues in Educational Privatization
HENRY M. LEVIN

Chapter 25 Charter Schools
ROBERT BIFULCO AND KATRINA BULKLEY

Chapter 26 Beyond the Rhetoric: Surveying the Evidence on Vouchers and Tax Credits
RON ZIMMER AND ERIC P. BETTINGER

Chapter 27 Emerging Alternatives to Neighborhood-Based Public Schooling
JOSHUA M. COWEN AND EUGENIA F. TOMA

Section VI RACE, SES, AND ACHIEVEMENT GAPS
SUSANNA LOEB, SECTION EDITOR

Chapter 28 Patterns and Trends in Racial/Ethnic and Socioeconomic Academic Achievement Gaps
SEAN F. REARDON, JOSEPH P. ROBINSON-CIMPIAN, AND ERICKA S. WEATHERS
Chapter 29 Early Childhood and the Achievement Gap
DAPHNA BASSOK AND SUSANNA LOEB

Chapter 30 Increasing the Effectiveness of Teachers in Low-Performing Schools
MELINDA ADNOT AND JAMES WYCKOFF

Chapter 31 Effects of Educational Policies on Disadvantaged Students: From Desegregation to Accountability
DOUGLAS N. HARRIS

Section VII SPECIAL CIRCUMSTANCES
DAVID H. MONK, SECTION EDITOR

Chapter 32 Special Education
TOM PARRISH, JENIFER J. HARR-ROBINS, AND JAY G. CHAMBERS

Chapter 33 Resource Needs for Educating Linguistic Minority Students
RUSSELL W. RUMBERGER AND PATRICIA GÁNDARA

Chapter 34 Challenges and Strategies Associated with Rural School Settings
JOHN W. SIPPLE AND BRIAN O. BRENT

Chapter 35 The Organizational and Fiscal Implications of Transient Student Populations
KIERAN M. KILLEEN AND KAI A. SCHAFFT

Chapter 36 Bridging the High School-College Divide
THOMAS R. BAILEY AND MELINDA MECHUR KARP

Contributors

Index
Preface

The purpose of the *Handbook of Research in Education Finance and Policy* is to assemble in one place the existing research-based knowledge in education finance and policy, and thus to help define this evolving field of research and practice. The *Handbook* traces the evolution of the field from its initial focus on school inputs and the revenue sources used to finance these inputs to a focus on educational outcomes and the larger policies used to achieve them. Researchers from multiple disciplines and with different methodological approaches examine a range of topics: how education revenues are raised and distributed; the changing patterns of education governance and finance, including privatization and accountability; supporting students with particular educational characteristics and needs; and the effects of these structures and policies on the equity, adequacy, and productivity of our education system.

This second edition of the *Handbook* builds on and updates the first edition, published in 2008 and conceptualized and coedited by Helen Ladd and Edward Fiske. The current version retains the original structure, but excludes the final section on higher education to focus more narrowly on elementary and secondary education, with one final chapter on the transition to college. It includes new chapters on cost-benefit and cost-effectiveness analysis, teacher accountability, and the growth of alternatives to traditional public schools. All the other chapters in this edition have been updated to reflect current developments and the most recent research in the field.

This *Handbook* is intended to serve as a comprehensive, balanced, and accessible resource for all those with an interest in education finance and policy. It is sponsored by the Association for Education Finance and Policy (AEFP) and is the official handbook of the Association. The mission of AEFP is to promote understanding of the means by which resources are generated, distributed, and used to enhance human learning. AEFP facilitates communication between and among various organizations and individuals in the school finance field, including academicians, researchers, policymakers, administrators, and teachers.

Background

The publication in 1983 of *A Nation at Risk*, the controversial report of the National Commission on Excellence in Education, was a milestone in American education. The report, which deplored the pervasive "mediocrity" of schooling in the United States, resonated with the concerns that many Americans had regarding the quality of education that their children were receiving. It placed education policy firmly on the national political agenda—where it has remained ever since. The search for ways to foster quality in education became a preoccupation of political and educational leaders and policy-makers at the local, state, and national levels. This concern about educational quality in turn had major consequences for the field of education research, whose practitioners were increasingly called upon to provide intellectual backup for policy decisions.

Education research can take many forms and ranges across many different topics. One fundamental distinction is between research that examines what goes on at the student level within a classroom and research that focuses on the structures that shape the conditions within which teachers and students interact. Like the first edition, this revised edition of the *Handbook of Research in Education Finance and Policy* focuses on research of the latter type. In particular, it concentrates on the financing and governance systems that are used to create and sustain the conditions necessary for promoting desirable educational outcomes.

Education finance emerged as a formal field of study at the beginning of the 20th century. What we now think of as traditional school finance focused primarily on the inputs of education—teachers and other human resources, textbooks, school buses, and the like—with special attention to how revenue from state and local taxes were raised and distributed among school districts. In its early years, the field of education finance
attracted a limited number of academic specialists; its practitioners tended to be accountants who were employed by state governments and whose engagement with policymakers was usually limited to annual or biennial deliberations over state budgets and appropriation bills.

In the 1980s the work of school finance—like the entire field of education policy of which it had by then become a part—took off in new directions. First came *A Nation at Risk*. Then, in 1989, President George H. W. Bush convened the country’s governors for an Education Summit on the campus of the University of Virginia to adopt a set of national goals for U.S. education. Policymakers and practitioners at the state and national levels began looking for ways to realize these goals. By necessity, discussions of school finance became increasingly intertwined with questions of broad educational policy.

This evolution continued into the 1990s and the early years of the 20th century. A turning point was enactment by Congress of the No Child Left Behind Act of 2001 (NCLB), which substantially altered the role of the federal government in the day-to-day functioning of local schools. Among other things, NCLB required school districts to administer tests in core academic subjects and imposed sanctions on schools in which substantial proportions of students in various categories failed to pass these tests. Students in “failing” schools were given the option of taking their public funding to another school. NCLB represented the political culmination of three major trends in national education policy over the previous two decades—each of which had major financial implications and produced new grist for the field of education finance.

The first was standards-based reform, a movement that stressed the need for ambitious educational outcomes and coherent policy structures to attain them. A major component of this trend is the shift away from a focus on inputs and processes to educational outcomes. Consistent with the shift was a change in the agenda of school finance research and policy from concerns about whether per-pupil revenues were equitably distributed across school districts to concerns about whether resources were adequate to achieve the desired outcome standards.

The second was the push for new forms of educational governance. These include school-based management, charter schools, and more parental choice within the public school system, as well as the possibility of greater use of public funding in the form of school vouchers for private schooling. These new forms contrast with the former “factory model” of education that was predicated on the view that one form of schooling can serve the needs of all students. New governance structures inevitably require new approaches to financing schools.

The third was increased emphasis on raising the productivity of the education system. Such an emphasis is a byproduct of the increasing concern with educational outcomes and draws attention to the concepts of accountability, incentives, and competition. This focus on productivity forces policymakers to ask not only how revenues are generated and distributed but how they can be used most effectively in the pursuit of the desired educational outcomes.

In short, a field that for the better part of a century was defined as education finance has evolved into the field of education finance and policy, a phrase that was captured by the title of the first edition of the *Handbook* and, as of 2010, by the name of the sponsor of this *Handbook*, the Association for Education Finance and Policy (AEFP), which previously had been the American Education Finance Association (AEFA).

**Structure of the Handbook**

This edition includes 33 of the original chapters, all of which have been updated to reflect current developments and recent research, and three new chapters. The first edition was organized to reflect both the evolution of the field and how each of the three broad trends have shaped content and methods. In addition, it recognized the continuity of the education process from preschool through college and the increasingly global marketplace of ideas in education policy. We have maintained the same structure for this edition with one significant change, namely the deletion of the section on higher education. So much new research on higher education has been produced during the last several years that a single section in this *Handbook* could not do justice to it. Clearly a second handbook on that topic would be useful.

The seven chapters in Section I, Perspectives on Education Finance and Policy, edited by Margaret Goertz and Helen Ladd, establish the central theme that the field of education finance and policy is no longer the province of accountants and economists alone. As decisions regarding school finance inevitably interact with
those of governance, accountability, and other areas of education policy, a full understanding of the important contemporary issues requires insights from a variety of perspectives. Thus the Handbook opens with a series of chapters that provide different perspectives on the field. The chapters in Section I show, among other things, how growing federal involvement in the delivery of education has altered relationships between the various levels of government, and how the public has frequently held broader expectations for schools than do legislators and other policymakers. Other chapters describe the growing importance of market-based and other economic concepts in education policy, the impact of litigation on issues such as equity and adequacy, and the emergence of sophisticated quantitative research methods. A final chapter highlights the growing importance of global trends in education finance and policy.

Section II, Making Money Matter, edited by Jennifer King Rice, explores the growing concern with productivity in education, a concept that has various meanings but is fundamentally concerned with using available resources most effectively to realize desired student outcomes. This section includes two new chapters, one that focuses on measuring teacher effectiveness and the other on the use of cost-effectiveness and cost-benefit analysis in education policy, reflecting the current policy interest in improving the quality of the teacher labor force and in making the best use of limited resources.

While researchers from different disciplines have used various methods to examine the relationship between education inputs and outcomes, much of the recent research on productivity draws heavily on the concept of the education production function as developed by economists. In general, policymakers and researchers have recognized three types of strategies to improve educational outcomes: (1) changing the quality or quantity of key inputs such as teachers, (2) improving the productivity of existing resources though better education technologies, and (3) introducing incentives embedded in high-stakes accountability policies or market-based reforms. Although researchers and policymakers alike have identified quality teachers as the single most important within-school educational resource available to students, the empirical relationship between specific measurable teacher qualifications and student performance is often elusive. Likewise, research shows that the effectiveness of incentives is largely a function of the design elements built into them, such as the outcomes that are included in an accountability system or who is eligible for a school voucher program. While theory suggests that all three strategies hold promise for improving student outcomes, continued research is needed to promote a better understanding of how resources matter in different contexts and for different types of students.

Section III, Promoting Equity and Adequacy, edited by Leanna Stiefel, explores the early 1990s shift from an implied to an explicit focus on achieving fairness and effectiveness in educational outputs. Previously, for example, states used foundation formulas to set a level of funding that would provide for the basic education of all children. Fairness was measured in terms of inputs, mostly expenditures or revenues, and output goals were only implied. With the advent of school finance court cases based on adequacy, the focus has shifted to outputs and, specifically, to the levels of financial and other resources needed to achieve specific learning outcomes. The design of systems to achieve adequacy is still developing. Researchers continue to refine measures of adequacy and methods for calculating the cost of educating students with different needs. We still know relatively little about the production function in education, including what role schools themselves can have in affecting outputs. Likewise, we have limited knowledge of the impact on learning of factors outside of the school walls, such as family resources and health issues, and the sorts of resources needed to compensate for outside-of-school deficiencies. Thus our ability to define the level of funding needed to achieve "adequate" education remains constrained.

The five chapters in Section IV, Changing Patterns of Governance and Finance, edited by Andrew Reschovsky, focus on the evolution of funding and governance in public education. Up through the first half of the last century, nearly all of the financial resources needed to fund schools came from local governments, with significant reliance on the local property tax. During the past 40 years, however, state governments have emerged as the largest single source of education funding. Research provides strong evidence that, starting with the Serrano v. Priest decision in 1971 in California, legal challenges to state funding systems led to an increased state share of support for schools and through the end of the century an increase in funding equity, usually through increased funding for low-spending districts. In a number of states, however, the increased role of state funding of education came with increased control by the state over the governance of the schools. This control took the form of requirements to take specific steps to improve the academic performance of
students and, in some states, of restrictions on how local school districts spent money and on their ability to raise local taxes. Although spending disparities declined during the final decades of the 20th century, new evidence presented in this volume shows that per-pupil funding inequities across the country rose sharply during the first decade of the 21st century, by some measures up to levels not seen since 1972.

The expanding role of the federal government, particularly with the introduction of NCLB and competitive grant programs such as the Race to the Top Program, also has far-reaching significance for education policy and finance and has generated substantial recent research. Finally, the growing involvement in education policy and practice of nongovernmental organizations, especially large national foundations, represents an important emerging area for research.

The chapters in Section V, Educational Markets and Decentralization, edited by Henry M. Levin and Ron Zimmer, describe a concept that appears to be clear from its language but is elusive in that it takes many different forms in practice. Virtually all schools produce both private benefits for families and public benefits for larger social entities, such as communities, regions, and nations, through higher productivity, democratic functioning, and social cohesion. The key starting point for researchers is to recognize that different forms of schooling, such as traditional public schools, charter schools, private schools, and home schooling, emphasize different combinations of public and private components and benefits. Moreover, the appropriate design of specific market-based reforms depends on the weight policymakers attach to different values.

Charter schools and voucher-financed private schools have attracted the most policy and research attention in the United States and, in the case of vouchers, also in other countries. New to this section for this edition is a chapter that focuses specifically on the supply of alternatives to traditional public schools. The pervasiveness of such alternatives varies across states largely because of differences in state laws and regulations. Clearly there are many forms of privatization, and predictions of their consequences are not straightforward. The details of each approach weigh heavily on the probable outcomes, and evaluations of results are challenging and often indeterminate.

Section VI, Race, SES, and Achievement Gaps, edited by Susanna Loeb, explores what is known about the relationships between family background and educational outcomes, focusing particularly on the reasons for these relationships and on policies designed to break the often deleterious link between the two. The various authors examine the research on patterns of achievement differences across racial and ethnic groups and among students of high- and low-socioeconomic status. Research has shown, for example, that systematic differences in development and ability emerge long before children enter school, and that children who start school at a disadvantage are likely to remain behind their peers throughout school and beyond. Advances in research have highlighted the importance of the early childhood environment as a contributing factor to school readiness gaps. Nonetheless, implementing programs that meaningfully narrow the differences in school readiness has been a major challenge.

Once children enter school, they are faced with resource differences that often contribute to the gaps. Research shows, for example, that the least-qualified teachers typically teach in the schools with the highest concentration of disadvantaged students for a variety of reasons related primarily to teachers’ preferences with regard to compensation and working conditions. Research presented in this section suggests a number of promising approaches to reducing achievement gaps. These include early childhood interventions, reducing segregation, providing financial incentives to attract teachers willing to work in challenging settings, and improving the working conditions in such schools.

Section VII, Special Circumstances, edited by David H. Monk, addresses the education of children who are in special circumstances that arise from certain combinations of geography and history, and the incidence of particular populations of students. It begins with a review of funding for students with disabilities and describes the lively debates that have surrounded both the diagnosis of their needs and the determination of the appropriate school response. Research in special education has shown the critical importance of establishing coherent links between special and general education funding. Successive chapters provide insights into the educational situation of students with limited or nonexistent English language skills, the particular needs and characteristics of students in rural schools, and the challenges that arise from transient students such as those who are homeless, are migrants, or live in low-income families.

This section, and the volume, concludes with an updated chapter from the first edition that focuses on efforts to smooth the link between high school and college. These include efforts to align high school
graduation requirements with college entrance requirements, dual-enrollment policies, and developmental education programs and services for academically underprepared college students.

Looking Ahead

As the general editors of the first edition, Helen Ladd and Edward Fiske produced a *Handbook* that reviewed the research on education finance and policy in a comprehensive, balanced, and accessible manner. They did so collectively with the help of a dedicated set of section editors, who devoted countless hours to their tasks. As the general editors of this revised edition, we are extremely grateful for the earlier contributions of Edward Fiske and, in particular, his commitment to high editorial standards. Our goals for this revised edition were both to update the research and to maintain that high quality. Fortunately, we benefited from the experience and skills of most of the initial section editors plus one new section editor (Ron Zimmer), and from the willingness of the original authors to update their chapters.

This volume is comprehensive in that it covers topics of traditional, current, and emerging interest. Traditional concerns include the evaluation of revenue sources, the distribution of inputs across school districts, and the design of state aid distribution formulas. Among the current issues are the use of the school as the basic unit for accountability and management, and the overarching concern with educational outcomes. Emerging issues include the challenges of dealing with students who have special needs, such as transient students and language minority students, the growing relevance of international comparisons and experiences, and the expanding role of large foundations.

The *Handbook* is balanced in that, in the selection of authors and the writing of chapters, we have sought to reflect the wide range of methodological approaches and conclusions that characterize the field of education finance and policy. While not shying away from controversy, we have tried to avoid pushing any particular point of view. We thank the contributors, many of whom have strong opinions on the topics about which they are writing, for taking this objective seriously.

Finally, we have paid special attention to the need to make the *Handbook* accessible to the growing audience of scholars, policymakers, and practitioners with an interest in education finance and policy. While many of the chapters cover complex state-of-the-art empirical research, the authors explain key concepts in language that nonspecialists can understand, in some cases providing technical details in appendices. This volume deals with issues that are important to all Americans—students, parents, employers, and citizens, as well as to education specialists and researchers. Our hope is that the wealth of factual information, data, and wisdom to be found in these chapters will make a significant contribution to improving the quality of education in the United States.

Helen F. Ladd and Margaret E. Goertz
Acknowledgments

Assembling a manuscript of this breadth and depth is a monumental task. The Association for Education Finance and Policy is indebted to the initial editors Helen "Sunny" Ladd and Edward "Ted" Fiske, and to Helen Ladd and Margaret "Peg" Goertz for overseeing this revision. A Handbook of this caliber would not have been possible without their powerful intellectual leadership, tireless attention to detail, and commitment to quality.

The Association is also grateful for the hard work and enthusiasm of the section editors as well as to the many contributors to this volume. Finally, thanks go to Barry Varela for his careful editing of the chapters, Heather Jarrow, and Abigail Stanley at Routledge for their assistance in the production of this volume.
Section I
Perspectives on Education Finance and Policy

SECTION EDITORS: MARGARET E. GOERTZ AND HELEN F. LADD
1 History and Scholarship Regarding U.S. Education Finance and Policy

MATTHEW G. SPRINGER, ERIC A. HOUCK, AND JAMES W. GUTHRIE

Introduction

America’s first generation of education funding systems provided land to stimulate popular provision of schooling. Today, state funding mechanisms have evolved into a complex array of policy levers ranging from pupil weighting schemes and interlocking systems of property, income, and sales taxation to provisions for vouchers, tax credits, and home school certificates. Scholars and policymakers routinely deal with issues ranging from how to promote more equitable distribution of education resources within states to the relationships (or lack thereof) between various “purchased inputs” and academic performance.

Much of modern education finance policy is rooted in questions that have persisted for over 100 years. For example, in 1906, Elwood P. Cubberley in *School Funds and Their Apportionment*, the work that launched the scholarly study of education finance, noted "However desirable and even necessary it may be to provide more money with which to maintain the schools of a state, a still more important question is how to distribute this money so as to secure the best results."

Like generations of education finance researchers to follow, Cubberley was concerned with the manner in which state governments generate and distribute funding to local districts for schools, and with the results that flow therefrom. Although he likely did not anticipate the full spectrum of issues now confronting education finance scholars and policymakers—economic concepts such as efficiency or productivity come to mind—Cubberley’s distillation of fundamental issues is as relevant today as it was over a century ago.

This chapter has a twofold purpose: (1) to describe the historic evolution of American education finance policy, and (2) building upon this historic platform, to explain the modern day integration of education finance with larger issues of education policy. In the process, the chapter reviews basic education governance structures and revenue generation and resource distribution mechanisms upon which American public education now depends. This discussion sets the stage for an examination of the three public values shaping the direction of contemporary educational finance and policy: equity, efficiency, and liberty.

Historic and Contemporary Context of Education Finance and Policy

Financial support for K–12 education is one of the nation’s major objects of public expenditure. School spending, as a percent of gross domestic product (GDP), has increased steadily since 1949 (see Figure 1.1). Since the early 1960s, there has been a fourfold increase in education spending per-pupil in constant dollars (dollars adjusted for inflation).
However, a reader should not view this trajectory as evidence that school spending has simply spiraled out of control. One reason for the fulsome increase has been the expansion of expectations within the larger policy system. These include increased high school graduation rates and consequent reduced incidence of school dropouts; the addition of costly auxiliary services such as meals, transportation, and guidance; the far greater inclusion of previously neglected populations such as disabled, limited English-speaking, and immigrant students; and a recent interest in the provision of preschool education for students. Public education continues to be increasingly labor intensive and, not coincidently, resistant to efforts to substitute capital for labor. Whereas economic sectors such as agriculture, transportation, manufacturing, communication, and finance have become increasingly capital intensive, schooling has moved in the opposite direction by adding labor and increasing the quality of that labor. Since 1955 the national average pupil-to-teacher ratio (see Figure 1.2) has decreased from 26.9:1 to an estimated 15.5:1 in 2010, the proportion of teachers with masters degrees and beyond has risen from 23 percent to 56 percent, and median years of teaching experience have increased from 11 to 15. Of greater interest is what Ingersoll, Merrill, and Stuckey (2014) call a "greening" of the teaching force. Whereas the greatest number of teachers had approximately 15 years of teaching experience in 1987, the modal years of teacher experience was five in 2011–12.
Conditions such as the continual expansion of school services, pressure for more efficient provision of services, an omnipresent desire for students to learn more and go to school for a longer period of time, and persistent questions about the quality of the teacher labor force are illustrative of the policy struggles that have characterized American public schooling since its inception. What distinguishes the early part of the 21st century, however, is that the amplitude or valence of these sometimes conflicting expectations has intensified. The growing significance of formal education, both for societal and individual well-being, has thrust education finance policy, and the research that accompanies it, into the forefront of the education policy and scholarly arenas.

Education Governance, Revenue Generation, and Resource Distribution

For more than 150 years, education has been a quasi-public good financed by taxes and provided free of charge to the immediate user. Governance in K–12 public education is a function of the interjurisdictional dynamics put in place by the U.S. Constitution at the end of the 18th century. The taxation and resource distribution mechanisms upon which American K–12 public education now depends was a product of this governance system and was itself shaped in the industrial period from the mid-19th to the mid-20th centuries.

Federal Beginnings and Their Long-Lasting Influence

The United States is one of only a handful of nations that rely upon a decentralized administrative and financing model, rather than a national or central governmental system, to operate and govern education. Because the framers of the Constitution did not explicitly specify education as a federal government responsibility, full authority for schools resides with state governments. States, subsequently, have delegated significant responsibility for financing and operating the nation’s approximate 14,000 public school districts and 96,000 traditional public schools to local authorities. School finance cases often fixate on how legitimate this delegated authority is; it varies based on state constitutional language and judicial temperament.
Political and economic support for schooling long has involved a mixture of federal, state, and local governmental actors and actions. Figure 1.3 shows that the role of federal, state, and local governmental entities in generating revenue and distributing resources has varied across time. Broadly speaking, local districts provide approximately 44 percent of total funding, the federal government 10 percent, and state governments the remaining 47 percent.

Early federal policy provided land for use in establishing schools, as land was one of the few resources easily at hand. As early as 1783, when England’s King George III’s Proclamation of Cessation of Hostilities culminated in the Paris Peace Treaty (1783), there was official reference to use of public lands to support schools (Barr et al., 1970). Since the Ordinance of 1785 (later known as the Northwest Ordinance of 1787), when the Continental Congress appropriated public lands to establish schools, federal land grants have served as a cornerstone for federal involvement in education. Such grants also provided a working model for state initiatives that induced local districts to establish an educational system.

It was not until the 1900s, and particularly the latter part of the 20th century, however, that the federal government began to appropriate money in a manner that forcefully influenced education policy. Each of these federal initiatives—from 1917 to the present—reflected national priorities, and each has influenced schools in the United States to the extent that vestiges of early efforts are still visible in today’s educational operations. Federal funding priorities for public schools have sought to ensure flows of scarce work-force talent (the 1917 Smith-Hughes Act, the 1958 National Defense Education Act), to expand access and equity (Title I of the 1965 Elementary and Secondary Education Act), and to promote school reform (1998 Comprehensive School Reform Act, provisions of the 2001 No Child Left Behind Act, the Race to the Top component of the 2009 American Recovery and Reinvestment Act). Federal involvement is most notable by the growth in the federal share of revenue provided to schools following the mid-1960s implementation of the Elementary and Secondary Education Act.

Because U.S. Constitutional authority for education was allocated, at least by default, to state government, and because state constitutions specify varying levels of commitment to education, revenue generation and distribution schemes vary widely across states. There are, in effect, 50 school finance systems operating nationwide, and even more if one considers the District of Columbia and the various federally overseen trust Territories and Department of Defense Schools. Add to this list the many private religious and independent
schools, and one begins to understand the remarkable complexity of U.S. school governance.

Authority for education rests principally with state legislative bodies. However, as education ascended as a public policy priority over the last quarter of the 20th century, increasing numbers of governors adopted education reform as a major element of their executive policy portfolio. For example, prior to becoming presidents of the United States, James Earl "Jimmy" Carter, Jr., William Jefferson Clinton, and George W. Bush were aggressive education reformers in their respective states. Beginning with education finance equity in the early 1970s and extending to education finance adequacy today, state courts too became increasingly active in overseeing education matters.

While federal and state involvement undoubtedly shaped early education policy, it was local townships and communities that bore principal responsibility for developing means to generate education revenue. This arrangement reflected Americans' generally high regard for local government and a belief that individual freedom is best protected if government remains in the hands of small, presumably self-concerned communities. Indeed, every state except Hawaii has delegated major responsibility for operating public schools to local school districts.

Most districts have local taxing authority, set salary levels for employees, allocate teachers and students across schools, and advocate for funding of capital construction. Local funding comes from property taxation and some local option sales taxation. The costs of capital construction are often spread over time through the use of bond finance.

**Industrialization’s Contributions**

Starting in the 1850s, industrialization fueled expansion of America’s public school system. The “new” economy involved greater “use of science by industry, a proliferation of academic disciplines, a series of critical inventions and their diffusion” (Golden, 2004). During this period, policymakers introduced taxation schemes for generating revenues and mechanisms to distribute these funds to local communities and schools. Although methods for generating revenues and distributing resources have continually been modified, their skeletal components assembled during a century-long time starting in the 1850s have remained comparatively unbroken.

**Generating Revenues.** Taxation is the main method of revenue generation for public education. Figure 1.3 provides a rough breakdown of education revenue by source. Within these breakdowns, 79 percent of the revenue provided by local governments was raised through property taxes. Sixty-eight percent of the revenue provided by state governments is revenue from state income and sales taxes.

**Distributing Resources.** In an effort to promote the provision and expansion of public schooling, the first generation of finance scholars began studying and developing inter-governmental fiscal arrangements and criteria for distributing revenue to compensate communities for differences in local property wealth. Scholars such as Elwood P. Cubberley of Stanford University, Arthur B. Moehlman of the University of Michigan, Henry C. Morrison of the University of Chicago, Harlan Updegraff of the University of Pennsylvania, and George D. Strayer, Robert Murray Haig, and Paul R. Mort of Teachers College not only contributed to an expansion of scholarship, but also had substantial practical influence upon policies that states adopted to generate and distribute revenues to local school districts (Johns, 1938; Moehlman, 1925, 1932; Morrison, 1924, 1930, 1932; Mort, 1924, 1926, 1932; Updegraff, 1922; Updegraff & King, 1922).

Three of the most common early methods that states used for distributing assistance to local school districts are: (1) flat grants, (2) foundation programs, and (3) equalization plans. The modern versions of these school funding strategies are discussed in Picus, Goertz, and Odden in this volume.

**Flat Grants.** When states began appropriating money to local communities to assist with the cost of schooling, they typically allocated equal amounts of money to each community, regardless of the number of school-age children or its ability to raise money, to guarantee that each locality have at least one public elementary school. This approach changed over time to take into account the size of the student population. By the 1930s, 38 states were distributing a minimum amount of revenue per pupil to districts based on a count of all school-age children.

Policymakers assumed that schooling in excess of this minimum level would benefit only the individual recipient or the community in which he or she resided. They further assumed that added spending above the
flat grant amount was a local luxury to be indulged in as each community saw fit, and not to be subsidized by the state. Since the amount of the flat grant is presumed to be sufficient to cover the education level the state believes to be minimally necessary, improvements in educational services were seen by policymakers simply as matters of increasing the minimum level, and as Cubberley asserted, should be made "without reference to whether or not a portion of its communities will be unable, unaided, to meet the demands" (1906, p. 17).

A practical problem with flat grants, however, is that states had no scientific means for determining the appropriate dollar amount per student. The size of the flat grant was determined by the political process and was usually lower than the resource level that school advocates believed a proper education required. Moreover, flat grants historically lacked any consideration of the diversity of students enrolled in districts, district size, or urban and rural contexts (Waterman, 1932; Mort, 1933).

**Foundation Plan.** George D. Strayer and Robert M. Haig (1923) were the first to publish descriptions of the practical problems with the administration of flat grants and to propose a solution. In a report to the Educational Finance Inquiry Commission based on a study of New York State, they proposed a system that captured a portion of the local property tax for state purposes through a required local tax levy. Their proposal has since become known as the foundation program, or the Strayer-Haig plan.

In a foundation program, just as with a flat grant, a state specifies the dollar amount per student to which each school district is entitled. Presumptively, this is the amount of money per pupil necessary to guarantee a minimally adequate education. At the time of the foundation plan’s conception, there were few systematic efforts to ensure that the dollar amount prescribed was sufficient to provide that level. The foundation dollar amount was often an outcome of the political process where state legislatures, predisposed to a certain level of taxation or constrained by the level of revenue available, initially established a revenue pool and then, through division, determined the per-pupil amount of the foundation.

As with the flat grant, and indeed all resource distribution schemes, there are practical problems with the foundation plan. First, despite a certain “elegance,” foundation plans still assume that the foundation spending is the dollar amount necessary for a minimally adequate education, although there is no way of determining such a benchmark with precision (Benson, 1961, p. 205). Second, if the tax rate mandated for the required local effort is set low, the aggregate amount of the funding appropriation required from state sources may be more than the state can afford. On the other hand, if the required local tax rate is high so that less money is required from the state, a substantial number of districts will raise more than the foundation guarantee at the local tax required rate, and hence lose eligibility for state aid.

**Percentage Equalizing Plan.** A percentage equalizing plan is based on a different philosophy than flat grants or foundation programs. Essentially, percentage equalizing plans define equity as equal access to funding for education. The state’s responsibility under this plan is to ensure that district level effort to support schools is equally rewarded regardless of any individual district’s property wealth. The state backs up this promise by subsidizing the tax effort of poorer districts.

Adoption of percentage equalizing was first urged by Updegraff and King (1922) around the same time Strayer and Haig advocated the foundation plan. However, it was not until Benson (1961) popularized the percentage equalizing concept that states began to adopt the approach. Yet, percentage equalizing was never widely adopted. In 2008, the foundation plan was the dominant vehicle for allocating state aid in 38 states compared to three states that relied on the percentage equalizing plan. Another seven states used a combination of the foundation and percentage equalizing plans (Verstegen, 2011).

Like the other plans, percentage equalizing has suffered from practical problems. First, if the local district is responsible for deciding the size of its education budget and the state thereafter guarantees equal access to funds, there are no budgetary size restrictions. For example, a district with a very high aid ratio could spend a large amount of money simply because it knows that an additional dollar of local taxes will bring in an additional $10 or $20 of state money at very little financial cost to its citizenry. Second, unlike a flat grant or foundation program, the state does not establish a minimum spending level under this funding approach. Third, the percentage equalizing formula, like that of the foundation plan, carries the possibility that some districts will receive no equalization money at all, or worse, will be forced to contribute to the state instead.

**Modern Education Finance: The Pursuit of Equality, Efficiency, and**
Liberty

Four scholarly publications, two highly visible national reports, and two court decisions have shaped modern education finance and propelled the field of education finance onto its current center stage of American education policy setting. Key questions confronting researchers and policymakers have included themes of equity, efficiency, and liberty. These three values, each important in the American political tradition, exist in tension throughout many school finance policies. The tensions between these values act as checks on their growth, providing for policy equilibrium with a center point that shifts as national political mores and powers shift.

In practice, policymakers and researchers do not articulate questions about values in equilibrium or tension. Instead they ask more policy relevant questions, such as: Are some students systematically disadvantaged by school finance structures? Is the public getting the most it can expect from the expenditures invested in public education? Are school systems maximizing parental choice in school attendance decisions? The rest of this chapter addresses these issues in turn, providing a backdrop for some of the chapters to follow.

Equity: The Quest for Equality

Despite the continual formulaic tweaking of funding programs throughout the first part of the 20th century, the level of education revenues varied across localities within a state. In the latter part of the mid-1900s, school finance researchers began to understand that such funding arrangements provided a systematic disadvantage to specific classes of students and citizens and sought to develop legal arguments to address this disparity. Private Wealth and Public Education, by John E. Coons, William H. Clune, and Stephen D. Sugarman (1970), and Rich Schools: Poor Schools, by Arthur Wise (1970), became path-breaking volumes. Though written independently, both books targeted property-related resource disparities within states as an injustice and constructed a constitutional argument, based on the Fourteenth Amendment’s equal protection clause, by which resource disparities could become a basis for adjudication. This scholarship, undertaken within the context of the 1960s Civil Rights movement, facilitated the equity quest of the 1960s and the judicial mindset of modern education finance that continues to this day.

In the 1971 case of Serrano v. Priest, plaintiffs argued that children living in property-poor districts received unequal treatment because the level of education funding in California was a function of community wealth. The California Supreme Court ruled in their favor, finding that that education was a fundamental right under both the U.S. Constitution and state constitutions, and school district wealth was a "suspect class," which gave the state a high hurdle to pass in proving the state funding system constitutional. However, in its 1973 decision in San Antonio Independent School District v. Rodriguez, the U.S. Supreme Court ruled that education is not a fundamental right under the U.S. Constitution, thereby, at least for now, removing finance equity as a federal constitutional matter.

Nevertheless, the legal reasoning based on state equal protection constructs by Wise and Coons, Clune, and Sugarman continued to prevail at the state level. The Coons team and colleagues termed these arguments Proposition One: The quality of a child’s schooling should not be a function of wealth, other than the wealth of the state as a whole. It is this formulation that provided the judicial system with a purchase on remedy, a criterion by which equity reforms can be judged and served as a foundation for much school finance research around issues of equality of access to educational resources (Guthrie, Springer, Rolle, & Houck, 2007). Proposition One does not specify any appropriate spending level per pupil. Rather, it takes the opposite approach and specifies what the pattern of finance distribution should not be. Whatever an appropriate level, it should not be a function of local or personal wealth.

Prior to the formulation of Proposition One, courts avoided education finance litigation for fear that there were no manageable solutions to finance inequality. This judicial reluctance was reinforced by seemingly daunting issues of definition and measurement. How could one tell if a state was complying with a judicial decree to render its education finance system equitable?

The next signal contribution was made by education finance scholars Robert Berne and Leanna Stiefel. In their 1984 volume, The Measurement of Equity in School Finance, Berne and Stiefel specified multiple
definitions of equity and methodological procedures for measuring equity, and explored empirical issues in resource allocation that facilitated plaintiffs’ pursuit of greater equalization in the distribution of school funding. Ultimately, their work changed the debate about school finance from theoretical (or ex ante) analysis to more concrete (or ex post) terms. Berne and Stiefel’s measurements and conceptual framework allowed researchers to measure the equity of funds as they had been generated and distributed, thereby giving lawyers and judges both a means to measure the effect of a system as well as a manner of defining how a system had to change in order to meet constitutional muster.

The legal logic constructed by the Coons team and Wise, buttressed by the measurement contributions of Berne and Stiefel, had an enormous effect on education finance and policy. Court cases alleging equal protection violations were filed in literally dozens of states, and by the 1990s 36 states had faced equity-based challenges (Springer, Liu, & Guthrie, 2005). Legislatures in many states—sometimes in response to or in anticipation of a judicial decision, or in hopes of forestalling an unfavorable judicial decision—began devising new distribution arrangements to reduce interdistrict per-pupil spending disparities.

This combination of state-level judicial and legislative efforts aimed at reducing resource disparities had significant impact on the distribution of school resources. Using a nationwide panel dataset with more than 16,000 district observations, Murray, Evans, and Schwab (1998) assessed whether funding disparities had decreased within and between states between 1972 and 1992. They found intrastate inequality fell more in states with court-mandated reform compared to those without these decisions. They also found that spending rose in the lowest and median spending school districts and remained constant in the highest-spending districts; and they showed how increased spending was a result of higher taxes and not a reallocation of resources from other government expenditure categories such as hospitals, health care, and highways. See Section III of this volume for a more comprehensive review of education finance equity.

The Quest for Adequacy (Equity II)

While the intrastate equity movement dominated education finance and policy during the 1960s, 1970s, and 1980s, it failed to address the amounts of money deemed necessary or adequate for students to achieve state constitutional guarantees; nor did it suggest means by which districts could spend resources more wisely in order to generate desired outcomes. It was a more general education policy movement in latter part of the 20th century that allowed school finance as a field to begin to tackle these difficult issues.

The standards-based reform movement, based on a theory of systemic alignment (Smith & O’Day, 1991), allowed policymakers to move from questions of resource distribution to resource use. Smith and O’Day articulated a vision for education policy by which state level policy defined the knowledge and skills that students would acquire, as well as assessments to measure the attainment of that knowledge and those skills. Local districts would align curriculum and instruction with state standards and prove their acquisition through performance on state level assessments. In this system, policies, curriculum, and assessments would be aligned in the service of teaching and learning. The adoption of a systemic framework facilitated the alignment of local, state and federal education policy—often around notions of educational accountability. Two questions arise when researchers inquire regarding the use of revenues. First, do funding levels provide sufficient support to schools in attaining achievement goals? That is, are the funds adequate to meet the expectations of the system? And, second, are educational funds expended in a manner that makes efficient use of taxpayer dollars?

A Kentucky Supreme Court decision in Rose v. Council for Better Education (1989) shifted school finance discussion from the conventional notion of equity—which focused on measuring the distribution of resources across districts—towards a new concept called resource adequacy or Equity II—which focuses on the relationship between the level of funding and the sufficiency of the funding to achieve state goals. Indeed, state courts have had substantial influence in placing the issue of funding equity and adequacy on the policy agenda (West & Peterson, 2007). According to the National Education Access Network, as of 2013 only five states had no court decisions on school finance equity or adequacy while 25 states had plaintiff liability victories, 19 states had defendant state liability victories, and one state considered mixed results. Between 1989 and 2006, 36 states had the constitutionality of their funding decisions challenged on adequacy grounds; of these challenges, 24 were ruled unconstitutional (Springer & Guthrie, 2007).

Adequacy seeks to “backward map” policy expectations to arrive at the required amount of student funding.
(Clune, 1995). While the definition of an "adequate" education varies from state to state, based on the language in a state’s education clause, the principle of education finance adequacy generically proposes that a sufficient level of resources be available to all students, thus providing them opportunity at least to reach a level of proficiency defined by state standards (Guthrie, Springer, Rolle, & Houck, 2007; Ladd & Hansen, 1999; Odden & Clune, 1998). While states have historically set their own state standards and proficiency levels, the federal government has provided financial and regulatory incentives for states to adopt common academic standards and assessments through its Race to the Top program and Title I waivers, a movement that may affect both the definition and cost of an adequate education.

Scholars have proposed four methods of approximating the cost of education programs and services in order to specify an adequate educational opportunity. These are the econometric or cost function approach, the successful schools or "empirical" approach, the state-of-the-art or research-based approach, and professional judgment models. Each of these methods for determining the cost of an adequate educational opportunity has strengths and weaknesses. Some authors argue that, given the current state of the art, it is not possible to construct credible estimates (see Hanushek, 2006; Springer & Guthrie, 2007). Yet, by 2013, 39 states had conducted at least one type of adequacy study (National Education Access Network, 2014). Downes and Stiefel (this volume) provide more detail on the measurement of adequacy.

The Quest for Efficiency

Efficiency as a concept focuses on the relationship between spending and educational processes and outcomes. Prior to the 1960s, researchers conceived educational efficiency in terms of standardization and consolidation. It was deemed less efficient, for example, to have a 30-member school board made up of representatives from each neighborhood in a city than to have a nine-member school board composed of more broad-minded professionals (see, e.g., Tyack, 1974). Similarly, the single salary schedule for teachers was deemed by school leaders such as William T. Harris to be a more equitable and efficient manner of compensation than negotiating individual teacher contracts. Harris also saw age grading—the process of placing students in a school grade based on their chronological age—as an efficient policy for creating city-wide school systems. These and other efforts are well chronicled by Raymond Callahan in his 1962 classic, Education and the Cult of Efficiency (Callahan, 1962).

Under the guise of accountability, another movement coalesced around ideas of efficiency in the late 1960s. The principal impetus for this efficiency movement was the federally sponsored report, Equality of Educational Opportunity, the prime author of which was the renowned sociologist James S. Coleman (Coleman et al., 1966). Congress commissioned the oft-referenced "Coleman Report" to document the depth of resource inequality that separated black and white schools. Although Coleman found abundant inequality, his study also found that more variation in performance could be attributed to home and social environments than to school. The net effect of the Coleman Report was to trigger a conflagration of controversy regarding the possibility that student achievement was less a function of the amount of school resources than it was a function of external environmental factors. Researchers sought to determine if schooling (and school funding) mattered for students and to more precisely determine what the relationship was between educational spending and student outcomes.

Despite Coleman’s efforts to dampen the firestorm, the popular press took these research findings to mean that dollars for schooling did not make a difference. The Nixon administration capitalized on the Coleman Report and attendant publicity to call for greater efficiency in the operation of America’s schools and to extol the virtues of school accountability, getting greater output within the confines of existing funds (Grant, 1973; Welsh, 1972). While followers of the 1960s accountability and efficiency movement attempted to establish specific educational goals, clearly affix responsibility for reaching those goals, precisely measure whether goals had been met, and often calculate the cost incurred (Murphy & Cohen, 1974; House, 1974), their efforts were largely overshadowed by the momentum of the previously described equity movement.

An increased focus on student outcomes, starting with the 1983 release of A Nation at Risk report, and with accountability under the 1994 reauthorization of the Elementary and Secondary Education Act, placed issues of educational efficiency as well as quality back on the policy and research agenda.

For researchers seeking the impact of the marginal dollar on educational productivity, two arguments,
perhaps predictably, emerged from new research: that money matters in producing educational outcomes (Hedges, Laine, & Greenwald, 1994; Laine, Greenwald, & Hedges, 1996) and, conversely, that money does not matter in producing educational outcomes (Finn, 1983; Hanushek, 1991; Mann & Inman, 1984; Walberg & Walberg, 1994). This debate was summarized and extended in the 1996 volume, Does Money Matter? (Burtless, 1996), and most concede today that money may be necessary in many situations. Newer research, instead of seeking relationships between dollars and outcomes, seeks to identify the effect of "purchased inputs"—such as teacher characteristics—and outcomes. These initiatives are discussed more fully below.

Policy proposals to increase educational productivity, that is, to gain additional output for constant levels of inputs, have centered on governance and market incentive. One such incentive, accountability, once popular, has reemerged, most notably with reauthorization of the nation’s omnibus Elementary and Secondary Education Act of 1965 in the No Child Left Behind Act (NCLB). The central objective of NCLB is that all traditional public school students, and defined subgroups thereof (e.g., race/ethnicity, free reduced price lunch status, and special education), reach academic “proficiency” by a specified school year. NCLB monitors progress toward a series of minimum competency performance targets that must be met by schools and school districts to avoid sanctions of increasing severity. In theory, NCLB’s threat of sanctions increases the motivation for schools and school districts to elevate learning opportunities for traditionally low-performing students and student subgroups. Figlio and Ladd (this volume) discuss research on the effects of NCLB on student achievement.

Perhaps the largest historically unaltered area of education finance is that of teacher salaries. The single salary schedule, which pays teachers based on years of experience and education level, has been a nearly universal feature of American K–12 school districts since the 1930s (Kershaw & McKean, 1962; Protsik, 1995). Policymakers have recognized that research suggests a marginal correlation, at best, between these two variables and student outcomes (Hanushek, 2003) and have coupled this recognition with the fact that the single salary schedule applies to roughly 95 percent of the nation’s approximate 3.1 million public school teachers (Podgursky, 2007). As a result, there is growing interest in merit- or performance-based compensation schemes in K–12 public education. Although merit-based pay programs date back to Great Britain in the early 1700s, and similar educator compensation reform efforts have been experimented with in the United States periodically since the 1920s (Hall & Stucker, 1971; Murnane & Cohen, 1985), most programs were short-lived. Research highlighted the difficulty inherent in creating reliable processes for identifying effective teachers (particularly with respect to the measurement of teachers’ value-added contributions), elimination of preferential treatment during evaluation processes, and standardization of assessment systems across schools (Murnane & Cohen, 1985; Hattr, Greiner, & Ashford, 1994).

One pay reform strategy that has gained some traction takes the form of knowledge-and skill-based pay programs that reward teachers for acquisition of new skills and knowledge presumably related to better instruction (Odden & Kelley, 1996). Although proponents argue that these strategically focused rewards can broaden and deepen teachers’ content knowledge of core teaching areas and facilitate attainment of classroom management and curriculum development skills (Odden & Kelly, 1996), evidence that the training and credentials being rewarded in these systems actually improve student outcomes is inconclusive (Ballou & Podgursky, 2000; Hanushek & Rivkin, 2004).

Another option is to link teacher pay directly to student performance. The federal government’s Teacher Incentive Fund (TIF) and Race to the Top programs, for example, support efforts to develop and implement performance-based teacher and principal compensation systems in high-need schools. Initiatives such as this are leading to additional state experimentation with salary structures including the elimination of additional pay for master’s degrees, renewed interest in pay differentiation for additional responsibilities, elimination of tenure or career status, and a shift of salary dollars towards early year teachers.

The direct evaluation literature on "pay-for-performance" plans, however, is slender. The nation’s first randomized field trials specifically designed to examine the impact of financial incentives for teachers on student achievement and institutional and organizational dynamics found little effect on performance of additional salary bonuses based on improved student test scores (Springer et al., 2013; Springer et al., 2012). Similar results were found for a school-level incentive pay program in New York City (Fryer, 2011; Marsh et al., 2012), while Fryer and colleagues (2012) found loss aversion to be a more powerful motivator.
The Quest for Liberty

The notion of liberty, or the opportunity to choose among schooling options, gained currency in American education policy upon publication of Milton Friedman’s (1964) voucher proposal in *Capitalism and Freedom*. Subsequent scholars engaged Friedman’s ideas to develop and articulate policy options for family educational choice in the modern era. Although private schools have been available since the inception of the Republic, available to those who could afford them, and their existence was legally legitimated by the 1925 Supreme Court case *Pierce v. Society of Sisters*, choice has never been systematically included as a policy value for all citizens.

Coons and Sugarman (1971) identified the “family versus professional dispute” as the distinguishing question of liberty argument. The family versus professional dispute recognizes that the family’s right to make decisions in the best interest of their child can come into conflict with educators’ professional obligation to make similar decisions. One example might be student assignment, where a parental desire for schooling closer to home can conflict with an educators’ desire for racially or socioeconomically balanced student populations. More broadly, the family/professional dispute points toward the fact that education produces both individual and public goods; additional education enlightens and enriches the student while also enriching the society within which that student lives. Public provision of education is assumed to account for these societal contributions while private provision is often viewed more as an individual transaction. Although this impulse toward family choice was motivated by a deep concern for impoverished children and families often trapped in low-performing urban public school systems, resultant choice-oriented policy proposals were envisioned for all parents (Coons & Sugarman, 1971, 1978). By overcoming enforced uniformity, it is hypothesized that the infusion of choice, or parental liberty, into the public education system creates a competitive market where pressures to compete among institutions leads to increased instructional innovation and student achievement (Chubb & Moe, 1990; Hess, 2002). An empirical justification of this argument, articulated by Hoxby (2002), has been challenged by Rothstein (2004) and others.

The market as an arbiter of educational quality is a powerful notion in current policy debate. Within this context, Guthrie and colleagues (2007) defined a continuum along which school choice options are best viewed, moving from the incremental (e.g., magnet school programs, intradistrict transfers) to moderate (e.g., charter schools) and then bolder initiatives (e.g., vouchers). Each of these forms of school choice is described briefly below. 2

**Magnet Schools.** Magnet schools seek to draw students from across a school district into specialized and/or rigorous programs, thus providing parental choice while contributing to efforts to desegregate schools without mandatory student assignment policies. Magnet school popularity received a boost in 1975 when the U.S. Supreme Court ruled in *Morgan v. Kerrigan* that magnet schools were an acceptable method of racially desegregating school districts. Indeed, between 1982 and 1992, the number of schools offering magnet programs nearly doubled (Goldring & Smrekar, 2000).

**Charter Schools.** The charter school movement began legislatively in 1991 in Minneapolis, Minnesota. Since then the charter school movement has gained significant momentum. From the 1999–2000 to 2012–13 school years, the number of charter schools grew from 1,542 to just over 6,000, while the number of students enrolled in charter schools reached 2.5 million in 2012–13.

While 42 states and the District of Columbia have enacted charter school laws, these laws vary widely regarding who may authorize charter schools, how many charter schools may exist in a state, how they are funded, and the extent of their autonomy (Wong & Shen, 2007). Despite wide variation in charter school characteristics, common elements include: an independent governance structure; reduced or eliminated state regulation of mission; direct parental control over policies and procedures through planning teams or other means; and a mandate to succeed with sanctions for failure, including revocation of the school charter (Finn, Manno, & Vanourek, 2000).

A second measure of charter school success is the extent to which charters create educational marketplaces and spur levels of competition sufficient to improve public schools. Research by Booker et al. (2008) suggested that greater penetration of the market by charter schools was associated with improved public school performance in Texas, while Bifulco and Ladd (2004) did not find similar effects in North Carolina. Research also suggested that charters inspired innovation in surrounding schools, but that a degree of political ill will
inhibited the transfer of innovative practices to surrounding schools (Teske et al., 2001). An updated and expanded review of research on charter schools can be found in Bifulco and Bulkley (this volume).

As research continues around charter schools, a series of policy debates have begun to coalesce: about the makeup of charters (charters writ large are more racially isolated); the governance of charters (states debate the legality of creating independent charter boards versus leaving charters to function under the auspices of state boards of education); the performance of charters (charters seem to "stretch the distribution" of school performance by exhibiting examples that far outperform traditional public schools while others dramatically underperform); and the scope of charter schools (charters sponsored by nationwide not-for-profit chains and for-profit vendors are debated, as are the creation of "virtual" charter schools that may exist functionally outside of state boundaries). Additionally, studies are beginning to examine the effect of charter schools on high school graduation rates, college attendance, and future earnings (Booker et al., 2011; Booker et al., 2014).

Vouchers. Voucher proposals have sought to make the portability of K–12 educational funding comparable to that of federal funding for higher education—where educational loans for college tuition are made directly to the student rather than to any other institution or governmental entity—by placing the control of educational funds with parents and families. Vouchers flow directly from Friedman’s arguments. Funds are provided directly to families for use in purchasing educational services in the private market. The theory behind voucher programs is similar to that of charters—private schools may increase student achievement, and market penetration and resultant competition may spur public school improvement. In the United States, findings from experimental and quasi-experimental studies have been mixed. An early study of the nation’s first education voucher experiment, the Milwaukee (WI) Parental Choice Program, found no significant differences between voucher students and public school students (Witte, 1998). Other studies of the Milwaukee data have found a positive voucher effect in math scores, but not in reading (Rouse, 1998). A second randomized study in multiple sites by Howell and Peterson (2002, 2004) found achievement effects for voucher students that were limited by subject area (math) and student racial background (African American). This research, however, has been reanalyzed resulting in a contrary conclusion sensitive to the classification of race and ethnicity for some study participants, specifically those of Dominican heritage who were classified as African American instead of Hispanic (Krueger & Zhu, 2002). Voucher plans continue to find homes in state legislatures, or at least in policy debate, and today are frequently referred to as "opportunity scholarships." Zimmer and Bettinger (this volume) review more current research on vouchers.

Performance debates aside, the liberty argument does not require that choice options perform better than traditional public schools. Rather, enhanced parental choice serves as the policy value as long as the performance of these alternative options is commensurate with that of traditional public schools, and the choice options do not work against the parental interests of others, including those who remain in the traditional public schools. The current accountability context and contemporary educational policy debate reflect the salience of this notion. NCLB sanctions (for example, in enfranchising students to "opt out" of their present school and to select another) and support for charter schools in other federal programs reflect the adoption of market principles in the education policy mainstream.

Conclusion

Over the course of 150 years, the financing of schools has moved from the provision of land and through a period of actuarial and political skill, to a point where issues of school finance intersect with and overlay many major policy initiatives in the United States. The range and depth of the issues examined constitute a long journey from the issues conceived by Cubberley in 1906. Better data, better methods, and better understanding of policy development have brought the field of education finance into a new era in which expectations for finance policy are aligned with education policy more broadly defined. Nonetheless, it is prudent to reflect again upon Cubberley’s initial concerns, and fitting that he should have the last word (Cubberly, 1906, p. 4):

> With the strong educational demand everywhere manifest for an improvement in educational conditions, ... the time is opportune, in many states, for a reopening of the question of providing adequate school revenue and for the revision of the general apportionment plan. The author would be glad if the principles laid down in the following pages should provide of service in formulating future legislation on the subject.
Notes

1. Ladd and Hansen’s 1999 National Council of Research volume, Making Money Matter: Financing America’s Schools was the first to refer to the second round of equity concerns, dealing with the adequacy of available resources, as “Equity II.”

2. A good reference on this topic is the Handbook of Research on School Choice (Berends et al., 2009).

References


Elementary and Secondary Education Act (1965).


inquiry commission IX. New York: Macmillan.
No Child Left Behind Act (2001).
Northwest Ordinance (1787).
Serrano v. Priest, 5 Cal. 3d 584 (1971).


The Role of Economics in Education Policy Research

DOMINIC J. BREWER, GUILBERT C. HENTSCHKE, AND ERIC R. EIDE, WITH TENICE HARDAWAY AND TIEN LE

Introduction

Over the last quarter century the discipline of economics has taken on growing importance in education research and policy. Topics on the economics of education make up a sizeable part of conference agendas for the Association for Education Finance and Policy (AEFP) and the Association for Public Policy Analysis and Management (APPAM). There is also anecdotal evidence that the economics of education has grown rapidly as a field within economics during recent years. Numerous scholarly articles are published in leading economics journals such as the American Economic Review, the Quarterly Journal of Economics, and the Journal of Human Resources, among others, as well as through specialized outlets such as the Economics of Education Review. The variety of applications has increased as well, building on labor and public economics to include institutional and behavioral economics.

The reasons for the growing importance of economics in the education policy arena are numerous. In the past, scholars and policymakers have tended to view economics and education as separate realms, with economics as a private good and education as a public good. Economics has been characterized as cold and impersonal due to its focus on rational, self-interested individuals and cost-benefit decision making, which on the surface appear to be unrelated to the social and moral values critical to educating children. Since A Nation at Risk (1983), however, the distance between the two realms has narrowed. It is well documented that better-educated workers have more favorable labor market outcomes than those with less schooling. Moreover, a well-educated labor force is critical for a nation to compete in an increasingly global economy that rewards knowledge and skills. The economic realities, both for individuals and the country, associated with the level and quality of education are therefore apparent, and they offer incentives to providers for improvement of education services and to households to pursue higher-quality schooling. Given that the study of incentives, choice, and competition lie at the heart of economics, economists have become more relevant to policy debates about education governance, organization, financing, and provision.

In particular, economists have contributed substantially to the study of market-based reforms in education provision. Economists have studied the effects of competition in the educational marketplace generally (Brewer, 1996; Ghosh, 2010; Hastings & Weinstein, 2008; Hoxby, 1999, 2003, 2004, and 2007; Lavy, 2010); as well as specific school choice options such as charter schools, which provide public funds to independent entities operating under a contract with a designated authorizing agency (Bifulco & Ladd, 2007; Hanushek et al., 2007; Zimmer & Buddin, 2007; Zimmer et al., 2009); and voucher schemes in which parents may use public funds at almost any school (Chakrabarti, 2008, 2013; Figlio, 2009). Use of markets applies to more than parental choice of school and includes reliance on buy versus make decisions in the provision of schooling services.

Another reason for the rising prominence of economics in education research and policy is the focus on standards- and outcomes-based accountability that has dominated the policy environment in K–12 education for the past 20 years. This movement toward accountability in education plays to economists’ strengths in a number of ways. First, there has been increasing attention to resource allocation, decision making at the school level, and viewing educational organizations as enterprises and those running them as entrepreneurs (see, e.g., Hanushek & Lindseth, 2009). Second, in order to understand educational policy in an era of accountability, it is necessary to consider more explicitly the way in which actors in a large complex system respond to incentives—for example, students responding to financial incentives for better performance, teachers reacting to merit pay incentives, or schools facing competition from one another (see, e.g., Bettinger, 2010). Third, as part of
building accountability systems, states, districts, and schools have become much more interested in the collection of data on outcomes and making them available in a way that permits their use in decision making. The federal government, through the National Center for Education Statistics, has also developed a systematic and extensive data collection program. Large-scale individual-level longitudinal data—including High School and Beyond (HSB), the National Educational Longitudinal Study of 1988 (NELS:88), and, most recently, the Beginning Teacher Longitudinal Study (BTLS) and the High School Longitudinal Study of 2009 (HSLS:09)—provided an unparalleled opportunity for economists to join their quantitative sociologist, psychologist, and political science colleagues in “mining” the rich data to investigate all sorts of educational phenomena.\textsuperscript{2} These data, combined with econometric know-how, permitted the use of more sophisticated statistical analyses in studying education topics. More recently the availability of student-level longitudinal data from states including North Carolina, Texas, and Florida has permitted a large number of new empirical education studies.

While economists’ use of improved data and econometric modeling contributed to raising the level of quantitative education research, the difficulty of teasing out solid cause-and-effect findings using multivariate nonexperimental methods helped push education research towards randomized experimental trials. With the Institute of Education Sciences strongly promoting this trend, it has had a marked influence on the field of education research. Further, given the expense and difficulty of actually carrying out true randomized experiments, economists have creatively used quasi-experimental methods, leading to a host of studies of educational policies and programs\textsuperscript{3} (see McEwan, this volume).

In this chapter we provide an overview of basic economic concepts relevant to education research and highlight some recent work that illustrates how the application of economic principles has been useful in understanding educational phenomena. We do not intend to provide textbook-like coverage of the main building blocks of economics, nor do we offer a comprehensive survey of the hundreds of books, articles, and papers using economics that might fall under the purview of “education finance and policy research.” Rather, we seek to characterize what we believe to be the major foci and trends of this work. We begin with an overview of foundational principles of economics relevant to education. We then discuss three major areas in which economics can usefully be applied to understanding educational issues: the relationship between education and the economic success of individuals and groups; the way in which society does and should allocate education services; and the organization and behavior of educational organizations as economic enterprises.

The Growing Use of Fundamental Economics Principles in Education

Economics is often defined as “the study of the allocation of scarce means to satisfy competing ends” (Gary Becker, as quoted in Walberg & Bast, 2003, p. 182). Economists study how individuals, organizations, and societies employ time, money, and effort. In the case of education, economists are interested in how society organizes and uses resources to produce various types of knowledge and skills through formal schooling and distributes them to various groups in society. This broad definition means that many social and political challenges can fall under the purview of economics. John Maynard Keynes once wrote that economics was a “way of thinking,” and it is this lens that has been brought to bear on a wide array of topics related to education. Perhaps the best modern illustration of this view is the bestselling book *Freakonomics* (2005), by University of Chicago economist Steven D. Levitt and Stephen J. Dubner. *Freakonomics* addresses topics such as what schoolteachers and sumo wrestlers have in common, why drug dealers still live with their mothers, and the impact of Roe v. Wade on violent crime, none of which are conventional economics topics. Along the same lines, economics blogs such as those of George Mason University economists Tyler Cowen and Alex Tabarrok (www.marginalrevolution.com) and Harvard economist Greg Mankiw (gregmankiw.blogspot.com) are gaining mainstream popularity by appealing to noneconomists who are interested in economic issues.

Economists typically begin an explanation of observed phenomena by building a theory or a model in order to simplify reality and highlight key characteristics. A model contains a set of assumptions and yields predictions, ceteris paribus (all other things being equal). Often this abstraction causes concern among noneconomists, but such simplifications are essential to understanding real-world settings. Economists would argue that what matters is whether the predictions of a model are correct on average rather than whether the assumptions underlying it are realistic. Economics, then, sits firmly within the tradition of theory-testing
scientific-method-based disciplines: a question is framed; a model/theory developed to explain behavior; and the hypotheses or predictions of that model/theory are then tested empirically using real-world data. Economics is often described as concerned with "positive" rather than "normative" issues, where the former are empirically testable and the latter are dependent on value judgments. Because of the emphasis on hypothesis testing, economists primarily use research designs that are quantitative in nature, attempting to discern whether predictions of cause and effect are true and whether they are generalizable. More recently, in their efforts to better understand the behavior of individuals in organizations, economists are also relying on a greater variety of methods where the rigor of randomized trials is not feasible or data is difficult to acquire (see, e.g., Strunk & Reardon, 2010).

Economic theories are typically built on three basic foundations: scarcity, rationality, and optimization. 

**Scarcity** refers to the assumption that individuals and society will never have enough resources to completely satisfy their unlimited wants. 

**Rationality** refers to people’s ability to make decisions in a systematic and purposeful way. It implies a “consistency of response to general economic incentives and an adaptability of behavior when those incentives change” (Ehrenberg & Smith, 2006, p. 4). 

**Optimization** refers to either profit or goal maximization in reference to organizations or utility maximization in reference to individuals. An individual or a group will have a particular goal—be it happiness, profit, or something else—and will make choices to maximize benefits, subject to the constraints faced. Again, these underlying assumptions often cause angst for non-economists who incorrectly conclude that economists care only about "selfish" or "self-interested" individuals. Economists view personal values much more broadly and use them to help explain individual behavior. In this vein, the "altruistic" values of Mother Theresa and the wealth-accumulating values of Donald Trump help explain the utility-maximizing behavior of both.

Economics, at its most general, asserts that in order to make sense of the world’s complexity it is useful to believe that behavior is motivated by the desire to achieve some sort of goal and that people behave consistently in trying to achieve that goal. Economists do not say that explaining why people (or institutions) have certain goals is or isn’t important; they usually leave that task to other disciplines, such as psychology. Similarly, although economists are interested in how and why people and institutions behave subject to the constraints they face and the context they find themselves in, they do not tend to delve too deeply into how those constraints were formed or whether the context is just. Historically other disciplines—history, philosophy, sociology, and political science—tended to have more to say on those topics. Yet as economists seek to provide greater explanatory power to their classical models of individual behavior, some have incorporated psychological principles into their economic models (see, e.g., Thaler & Sunstein, 2008).

Economics, then, is a framework for helping understand the behavior of individuals and organizations in allocating resources. Using this perspective, economists have examined a wide range of education-related topics, and in the remainder of this chapter we discuss three of the major questions of interest: How much education should an individual acquire? How should education be produced and allocated by a society? Can we be more efficient in organizing the production of education?

**How Much Education Should an Individual Acquire? The Relationship between Learning and Earning**

One of the most common areas of research for economists dealing with education has been the link between schooling and various individual outcomes, especially those associated with the labor market. Economists typically view education (and training) as an individual investment decision designed to achieve a monetary return in the labor market. This notion of "human capital" has a rich history, with early economists including Adam Smith, John Stuart Mill, and Alfred Marshall suggesting that individuals’ skills could contribute to their economic status. In 1776, Smith laid the foundation for human-capital theory when he wrote that human effort lies at the root of all wealth. In 1848, Mill built upon Smith’s notion; he considered human abilities as “means to wealth” (Sweetland, 1996). Modern-day human-capital theory has further extended the central insight through the pioneering work of Schultz (1963), Becker (1964), and Mincer (1958, 1962).

Underlying the concept is the idea that the knowledge and skills acquired through educational investments
increases human productivity enough to justify the costs incurred in acquiring them. Investments in human capital take place primarily during three life stages: early childhood, when learning is predominately controlled by parents and early schooling experiences; late adolescence, when students attend high school, college, or vocational training programs; and during the working years, when adults receive on-thejob training or attend night school or complete online courses. (This pattern is, of course, more characteristic of developed than developing countries.) At each stage of investment, one may incur costs in the form of out-of-pocket expenses, foregone earnings, and psychic costs associated with the pressure of studying and examinations. Benefits accrue later in life through enhanced earnings in the labor market, access to better jobs, a higher likelihood of being employed, and better health. There are also psychic benefits in the form of enhanced social status and the prestige associated with higher levels of education. Although each individual’s motivation for pursuing schooling may differ, and the psychic costs and benefits may be quite varied depending on personality and other traits, economists hypothesize that, other things being equal, the more education acquired, the higher the earnings achieved after the schooling is completed.

Prima facie evidence for human-capital theory is to be found in the strong positive relationship between education levels and earnings that exists in almost every developed country. For example, the most recent data available from the U.S. Census Bureau shows those investing in more education earn higher salaries over their lifetime. Generally, earnings rise with education level, and they increase at an increasing rate in the immediate post-education years, continue to increase at a slower pace, and then flatten as individuals approach retirement (Ehrenberg & Smith, 2006). This general pattern of earnings by education level holds for almost all subgroups, including men and women and different racial and ethnic groups, but it is the differences among these groups that often fuels education policy debates about the distribution of education subsidies and services. As Table 2.1 shows, whites earn considerably more than other races regardless of level of education.

Such raw statistics do not, of course, prove that education causes higher earnings, because we cannot rule out the possibility that other factors are responsible for the positive relationship that is observed. For this reason, economists have devoted considerable attention to the challenge of estimating the returns to additional years of schooling taking account of other factors. Returns to investments in education have been estimated since at least the late 1950s. Analysis of such returns generally reveals a consistent positive relationship between investment in education and increased earnings for individuals, with an estimate of the average rate of return to an additional year of schooling at about 10 percent (Psacharopoulos & Patrinos, 2002). Estimating such rates of return proves to be very complicated statistically, given that individuals are assumed to choose their education levels. Significant advances have been made over the past three decades both in available data and in statistical techniques that have permitted more reliable estimates of the causal effect of schooling on earnings. These approaches range from use of statistical models that rely on instrumental variables (see McEwan, this volume) to studies of twins (Ashenfelter & Krueger, 1994; Ashenfelter & Rouse, 1998; Behrman et al., 1994; Card, 1995, 1999) to regression discontinuity (Oreopoulos, 2006).

Table 2.1 Earnings in 2012 by Educational Attainment of Workers 18 Years and Over, by Age, Race, and Hispanic Origin
Continued interest in estimating returns to education has been spurred in part in recent years by interest in what appears to be an increasing return to higher levels of education. Several studies have examined the changes in education-related wage differentials over time (e.g., changes in average earnings of high school graduates relative to average earnings of college graduates) and begun to explore the underlying reasons for the shift (e.g., Brewer et al., 1999; Grogger & Eide, 1995; Murnane et al., 1995). Similar education-related wage differentials are also found among regional labor markets and countries as well as across time (see, e.g., Psacharopolous & Patrinos, 2002). Greater returns to schooling from region to region are often associated with greater overall economic productivity. Such research has implications, potentially, for education policy. For example, it suggests policies designed to encourage students to stay in school as long as possible, and to gain access to postsecondary institutions, have considerable economic benefit for some students. It also suggests that policy adjustments enhancing the content and quality of schooling (e.g., in terms of curriculum, resource allocation, teacher quality, and so on) might be worthwhile in order to take advantage of rising economic returns (e.g., Brewer et al., 1999; Grogger & Eide, 1995).

In addition, recent studies have focused on more contextualized estimates of the return to schooling. They have, for example, examined the return to different types of education diplomas and degrees (e.g., Kane & Rouse, 1993), to different college majors (e.g., Eide, 1994, 1997; Grogger & Eide, 1995), and to different types of institutions (e.g., Brewer et al., 1999; Dale & Krueger, 2002). These finer-grained studies are of more interest to the policy community because they begin to unpack the simplifying notion that all years of schooling are equivalent. Early education (i.e., preschool) appears to have particularly robust benefits, both social and personal, when overall benefits are evaluated net of costs incurred (Heckman & Masterov, 2007). This work reflects the increasing attention of economists to institutional and policy-relevant factors in rate-of-return studies.

Economists have also highlighted the strong correlations between educational levels and a number of other private and public social benefits. Individuals who have invested in education and job training (holding income and other factors constant) often have more job stability, have improved health (e.g., exercise more regularly, smoke less, and eat better), are more likely to receive employer-provided health insurance and pension benefits, are more inclined to vote, and have generally increased social and cultural capital that often enables upward mobility (Cohn & Geske, 1990; Dee, 2010; McMahon, 2010; Muennig, 2010; Wolfe & Zuvekas, 1997). This linkage between education and an array of positive outcomes that benefit society as a whole means that there is a great deal of public-sector interest in encouraging investments in education—in other words, in not leaving the decision to acquire additional schooling entirely to individuals (Dee, 2010; Karoly, 2001; Lochner, 2011). The extent and type of government involvement in the provision of education is the subject of the next section.

**Producing and Allocating Education Services: The Emerging Role of Markets**
The discussion thus far has focused on returns to education at the individual level. However, there is also general consensus that national investments in education lead to economic growth (for a review of the literature, see Sturm, 1993). Although the measures of economic growth vary, they all support national investments in education and job training (Hanushek & Kimko, 2000). Countries spend a sizable percentage of their gross domestic product (GDP) on education each year (see Figure 2.1). In 2010, the United States invested 7.4 percent of its GDP in education.

![Figure 2.1 Education as a Percentage of GDP, 2010 by OECD Country.](http://dx.doi.org/10.1787/eag-2013-en)

Having a high level of national income devoted to education is generally regarded as beneficial to society. Educational spending can be undertaken by private individuals and by governments through public expenditures. The decision as to how education at different age levels should be allocated is a central issue in public policy and, at its core, is an economic decision about how best to produce and allocate scarce resources. In studying this broad question, economists have focused a good deal of attention on two different approaches: those associated with "production functions" (essentially technical relationships akin to engineering) and those associated with organizational forms (essentially trade-offs between markets and hierarchies, or organizational design).

Economists generally assume that goods and services are produced in accordance with some technical relationship that combines different inputs (sometimes called "factors of production") to produce the desired outputs. Goods and services may get allocated in many different ways—by tradition, force, or lottery, for example. In modern societies, resources are allocated by markets, by governments, or frequently through the interplay of both. Most goods and services in the U.S. economy are allocated by markets, but K–12 schooling has traditionally been allocated by government at the federal, state, and local levels, with postsecondary education allocated by some combination of markets and government. As concerns about the effectiveness of existing schools have risen, policymakers have questioned the central role and functions of government in the allocation of educational resources, and have recently turned more toward market-related mechanisms such as contracting out, charter schools, and vouchers.

A market is a collection of buyers who purchase and sellers who produce and sell goods and services; the interaction of buyers and sellers results in the possibility of exchange and, hence, in the allocation of goods and services. The transaction is facilitated through agreement on price. The dynamics of markets requires a continuous process of adjustments that includes shortages and surpluses and involves consumers and producers entering and exiting the market. It should be apparent that in any market, firms will be opening and closing, causing disruption. In a school setting, the dislocations caused by the opening and closing of schools

---

38
affect children and their families, a cause for concern. For a competitive environment to work well in a K–12 school setting, therefore, attention must be paid to the incentives for good producers to grow enrollments and for protections for students who may be forced out of a failing school.

Markets are often the preferred method for allocating resources because they are able to coordinate many buyers and sellers, give consumers considerable influence over price and quantity, and avoid relying on a handful of arbitrary decision makers. Under many but not all circumstances, markets are the most efficient mechanism for allocating resources. In a narrow sense, this means that no more could be produced with the same resources, and the same could not be produced with fewer resources (productive efficiency). In a broader sense, markets can be "allocatively efficient," meaning resources are used in the way that consumers want. Note that no judgment is made here as to whether the distribution of resources is fair or equitable. Clearly, education has multiple goals (Gill et al., 2001) and this multiplicity of goals needs to be considered in deciding what is the best mechanism for allocating education from a societal standpoint.

Market failure occurs when markets do not efficiently organize production or allocate goods/services to consumers. There are several reasons why markets fail. First, market power may arise when a supplier of a good/service has the ability to control price. A monopoly is an example of such market power. Second, when consumers have incomplete information about price and product quality, markets cannot respond efficiently and correctly. In the case of schools, even if parents were given complete information, they may not choose schools based on what would be best for the broader society. Third, externalities exist when consumption or production has an indirect effect on others that is not reflected in market prices. In the case of education, the decision maker (e.g., a potential student) does not bear all the costs or reap all the rewards from his or her actions. Even though society may benefit more from an educated person, the person making the educational decisions may not see those benefits as his or her own. Thus, education will be underprovided from the perspective of the market. This presence of social benefits arising from basic education is perhaps the chief reason why governments have typically viewed education to be so important to the life chances of individuals that they have made it compulsory at elementary and secondary levels. Fourth, markets may fail for public goods—those that can be made available to additional people without additional cost (nonrival), and once provided are difficult to prevent others from consuming (nonexcludable). Markets tend to undersupply public goods.

The likelihood of market failure and, in particular, concern about the likely under-consumption of education by private individuals from society’s standpoint, as well as the importance of educational goals other than efficiency, has historically led to significant government intervention in the education sector. Indeed, in most countries, schooling at the K–12 level at least (and in many nations, postsecondary education too) has been allocated almost exclusively by the government. We can usefully break down the role of government into three functions—regulation, financing, and operation. Distinguishing between these roles is critical because arguments for and against government involvement (e.g., based on equity goals or externalities) do not necessarily imply that all these roles are needed.

Regulation can take different forms, including setting safety standards, mandating curriculum or student assessments, and requiring teacher credentials. Financing can be in the form of direct funding to schools or various forms of financial aid to individuals—for example, tax credits or vouchers. The former approach more directly supports the supply side of schooling, while the latter approach more directly supports the demand side. Revenues for schooling may be generated from general taxation rather than user fees, such that there is no clear relationship between receipt of the service and the payment for it. This simple fact is critical because it puts the responsibility for paying for education on a much broader constituency than simply those who directly consume it. The government may also directly oversee the operation of educational enterprises, meaning that the delivery units are embedded within a larger governmental infrastructure controlled by political mechanisms, "owned" by the state, and operated by employees who are civil servants.

Historically, school operation has encompassed regulation, finance, and operation in a vertically integrated public sector system. Further, because government-operated schools have been designed to serve all students in a geographic area, in that locality they constitute a virtual monopoly. From a market perspective, monopolistic schools do not face competitive pressure to keep quality high and costs low. In addition, many families (particularly low-income and minority families) do not have much choice over the schooling options for their children. The system is also biased toward public production; private production is possible, but families using
private schools still have to pay taxes.

Recent educational reforms are, to some extent, characterized by an attempt to unbundle regulation, finance, and operation. From its early origins in local communities, publicly provided schooling at the elementary level spread to the high-school level and beyond. The central role of government was not seriously challenged during this expansion. This changed nearly a half century ago when Nobel Laureate Milton Friedman (1962) questioned the fundamental premise that government was the best agent for allocating schooling and called for a system of private "vouchers" in which, using government financial support, individual parents would be free to select any school for their children. Such a system would enhance individual liberty. Further, if schools had to compete for students, over time they would become more efficient and more effective.\(^5\)

Over the course of the subsequent half century, changing household and schooling conditions fostered an array of arguments by others about the merits of specific elements of Friedman’s proposition (Hentschke, 2007). Chief among them was whether and to what degree vouchers should be in higher amounts for children from lower-income households (means-tested). In the early 1970s, for example, as concerns about household income inequality and social justice rose to the fore, John Coons and Steve Sugarman (1971) suggested that poor parents needed to be given greater choices than their wealthier counterparts in order to help ensure equal access to quality schooling.

With these salvos from the right and the left of the political spectrum, coupled with a growing recognition that governments were themselves self-interested economic actors in the education business, interest in alternatives to a government monopoly as the underpinning of the K–12 system began to be discussed more frequently and openly. Finally, with the growing concern about the nation’s economic health in the late 1970s and early 1980s, policymakers began to question the effectiveness of the nation’s K–12 system and entertain much more radical reforms than in the past. In parallel to this, economists began increasingly to examine the internal governance structures of schools and districts, seeking better to understand links between organization structure and performance—for example between strong union contracts and student performance gains (Strunk & McEachin, 2011). Consequently over the past three decades there have been various attempts to introduce elements of market-based resource allocation into K–12 education in the United States. Needless to say, these have attracted a great deal of attention from economists. The variants of market-based reform are complex—ranging from tax credit schemes (scholarships and deductions), to education savings accounts, to magnet schools, to controlled choice programs, to charter schools and voucher programs, to portfolio policies, to online options, to homeschooling (Friedman Foundation, 2014). Although we cannot provide an overview of all reforms, each contains some market-based elements, and a few require our attention due to their current scope and scale. Several chapters in this volume contain a detailed review of the evidence on each.

Beginning with the contracting out of noncore services to private vendors, state governments and local public school districts have gradually eroded the coupling of public school finance and public operation. The most significant reform in this regard has been the emergence of charter schools. Charter schools can be started by educators, parents, community groups, or others who enter into a performance contract with an authorizing agency, usually a local school district (Brewer & Wohlstetter, 2005). Forty-one states and the District of Columbia permit charter schools, and they serve more than 2.3 million students in about 6,000 charter schools across the country. The vast majority of charter schools are “new starts,” but many charter states also allow existing public schools to convert to charter status. The school’s performance contract spells out how the school will be organized and governed, how students will be taught, and the performance goals that will be achieved at the end of the contract period. The vast majority of charter schools are privately governed, and their operation is largely publicly financed. Charter schools are public schools of choice; educators, parents, and students choose to be there rather than being assigned by a district office, and these stakeholders have considerable autonomy to decide the school mission, how students will be taught, and how resources will be used. The theory of action underlying the charter school concept proposes that if schools are empowered to make their own decisions and they are schools of choice that must attract educators to work and families to attend them, then the schools will work to innovate to improve teaching and learning. It is costly for governments to adequately monitor schools regardless of whether schools reside within the hierarchy of a public-school district or are totally private and connected to the government only by a publicly provided private school voucher. The costs of monitoring must then be balanced against other gains and losses associated with these different supply side arrangements, for example gains and losses associated with control,
innovation, equity, and social cohesion. When public-school districts entertain entry into online education delivery along with approval of charter schools and other services, they evaluate the different costs and benefits associated with buying these services (through contracts) or making them (by growing their own capacity in-house).

There have also been a handful of attempts at a more explicit (demand side) market-oriented reform, in the form of vouchers along the lines originally envisaged by Friedman, Coons, and Sugarman. These attempts have included privately funded voucher schemes (e.g., New York) and small state-sponsored public schemes (e.g., Milwaukee, Cleveland, and Florida). In 1990 Milwaukee, Wisconsin, became the first city to offer vouchers for private schools, and as of 2013, 24,027 students were enrolled with an average voucher value of $6,442. Students whose family income is less than 300 percent of the federal poverty level are eligible. In 2011, Indiana adopted a statewide voucher system that would provide tuition money for low-income students transferring from public schools to private schools. In 2008, Louisiana enacted a statewide voucher system that benefited students whose family income is no more than 250 percent of the federal poverty line. On May 7, 2013, the Louisiana Supreme Court ruled that the voucher money could not be used to pay tuition costs at nonpublic schools.

The evidence on the numerous effects of these attempts to move away from government regulation, financing, and operation of K–12 schooling is growing. Needless to say, many of the findings are specific to the particular state (e.g., charter school legislation) or program context (e.g., whether a voucher is targeted to particular kinds of students), so making generalizations is difficult. One of the most comprehensive examinations to date of the effectiveness of voucher and charter programs (Gill et al., 2007) examined the available empirical evidence on voucher and charter programs in terms of their effects on five policy areas: academic achievement, choice, equitable access, integration, and preparation for civic responsibilities. The implications of these findings are that “In some contexts—such as high-poverty cities with substantial African-American populations, or communities that have dysfunctional public schools—targeted voucher programs may produce discrete benefits.” Moreover, “such programs will not be a silver bullet that will rescue urban education, but they are unlikely to produce the negative consequences that voucher opponents fear.” Economists view differences in the transaction costs and weigh them against differing bundles of services being produced.

The evidence on charter schools is more problematic to summarize because charter programs vary markedly across states and present significant methodological challenges (see Betts & Hill, 2006). Recent examples of empirical studies include Hoxby and Murarka (2009), Cremata et al. (2013), and Zimmer and Guarino (2013). (A comprehensive overview of recent empirical evidence may be found in the Bifulco & Bulkley chapter in this volume.) Research on the effects of competition in the educational marketplace is also complex due to the methodological challenges of isolating effects. Goldhaber and Eide (2003) provide an overview of the research that focuses on the potential achievement effects of greater private sector competition on K–12 schools. They find that although many of the methodologies used to assess the effects of particular interventions, such as educational vouchers, are sound, they likely fail to capture the general equilibrium consequences of enhanced choice due to their limited and partial implementation. For example, Wolf and Hoople (2006) noted that although randomized field trials (RFTs) can control for a number of confounding variables, a major shortcoming of RFTs is that they tell policy analysts little about why or how a policy intervention yields benefits. Taken together, these studies suggest that while voucher and charter school programs do hold some promise, it is too early for a definitive answer regarding their long-term benefits to students. Since there are no large-scale voucher programs in operation in the United States, and charter schools continue to be a small fraction of the total number of schools in most states, claims as to improvements in the overall effectiveness of schools remain unsubstantiated. Gill and Booker (this volume) provide an overview of recent literature in this area.

Opening the “Black Box”: Understanding the Inner Workings of Organizations
In addition to studies of links between learning and earning and to the overall production and distribution of educational services, economists have sought to better understand organized work, including education. These efforts have taken two different forms. One approach uses the notion of an "education production function," in which schooling outcomes are produced by combining inputs. In this formulation, schooling processes occur within a black box, with little attention to what goes on within the box. The second approach explicitly looks inside the black box and examines the organization as a web of interpersonal contracts wherein individuals seek to coordinate others (and are in turn coordinated by others) in the performance of work. This latter arena is most often referred to either as transaction cost economics (accounting for the newly recognized costs of coordination or transactions) or as applications of principal-agent theory (seeking to capture the complex issues of delegation of decision rights between bosses or principals, and their subordinates or agents). The first perspective has been applied more rigorously to research in education, while the second has been applied more implicitly to issues of education governance. The impact of both perspectives on education policy is notable, if mixed, and education policies continue to be developed that alter the decision rights of different education stakeholders, based on assumptions about their principal-agent effects (see, e.g., Taylor, 2013).

Economists often use an input-output framework to help think about schooling. The main inputs may include teachers, administrators, supplies, and facilities, while the main outputs are student achievement (knowledge and skills). Schools transform a set of inputs into a set of outputs using a production function. This simple abstraction is very complex in education practice (Goldhaber & Brewer, 1997; Rice & Schwartz, this volume). For example, it is difficult to identify and measure all inputs and outputs of schooling. Multiple outputs—for example basic skills, vocational skills, creativity, and attitude—are valued, may accrue in a cumulative manner, and may be discernable only many years into the future. Inputs can be difficult to measure, and the dimensions most easily measurable may not capture adequately the important features of a given input; for example, teacher quality may not be captured well by proxies such as years of experience or qualifications (Goldhaber, this volume). Nonschool inputs, such as peer influence and family background, clearly affect how much students learn. And outputs are themselves joint products; that is, students experience multiple teachers and carry with them knowledge from other classes and from home. The production-function approach, however, is valuable as a framework for thinking about what manipulable resources really make a difference for student outcomes, and, as such, it focuses attention on how schooling should be organized so as to get the most "bang for the buck."

Many studies have attempted to determine the relationship between inputs and outputs as currently exists in the United States. Researchers estimate statistical models using data on inputs and outputs across many students either at one point in time or over time, often focusing on a particular input, such as teachers (e.g., Boyd et al., 2005; Dee, 2004; Ehrenberg & Brewer, 1994; Goldhaber & Brewer, 1997 and 2000; Loeb & Page, 2000) or class size (Ehrenberg et al., 2001). There have been numerous, often-cited reviews of this large literature, notably by Hanushek. His (in)famous conclusion is that, based on this literature, there is "no consistent relationship" between most measured school inputs and student achievement (Hanushek, 1997). Although this conclusion has been disputed in terms of meta-analytic techniques (e.g., Greenwald et al., 1996; Krueger & Whitmore, 2001), this view has become widely accepted among economists and used in school-finance litigation against increases in school spending. Hanushek’s conclusion does not imply that schools or teachers don’t matter. Indeed, the evidence is that they do affect the overall level of student achievement, and that unmeasured characteristics of schools and teachers can account for significant variation in achievement from school to school and teacher to teacher. But it does suggest that under the existing organizational structure of public schools in the United States, additional resources are unlikely to generate significant improvements in student achievement.

Information on productivity gains attributable to any change in school inputs is interesting for policymakers but in principle is not sufficient for taking action, because the costs of changing inputs has to be considered. Because any program or policy uses both tangible resources such as facilities, equipment, and materials and nontangible ones such as time and energy of the staff, there are opportunity costs involved. ("Opportunity costs" refers to the resources used for any given activity that entail sacrificing what might have been gained from using the resources for another purpose.) Identifying all the ingredients that comprise an intervention, policy, or program (including personnel, facilities, equipment, and materials), placing a value on each, and summing all the costs is time consuming and rarely done (McEwan & McEwan, 2003).
One particularly salient area in which education economists (typically trained in public finance) have played a role is in the study of school-finance reforms that have dramatically impacted the level and allocation of educational resources (e.g., Corcoran & Evans, this volume). Beginning with *Serrano v. Priest* in California, many states have experienced challenges and changes to their systems of school finance. According to Murray et al. (1998), 43 states have challenged the constitutionality of the public-school finance system. Many of these states shifted local funding to the state level in response to court mandates. Murray et al. found that such education-finance reform efforts had three main results: decreased within-state inequality; reduced inequality by increasing spending at the bottom of the distribution while leaving spending at the top unchanged; and increased spending for education possibility through higher taxes. Loeb (2001), in her review of school-finance reform, points out that not only did the shifting of local funding to state funding impact the allocation of educational resources, the structure of the school finance system is also important. She compared those with pure state funding, state funding with unlimited local supplementation, and state funding with limited local supplementation and concluded that indeed the type of state funding system strongly affects education spending. These studies serve to illustrate the involvement of economics researchers in the school-finance arena.

The second economics-oriented perspective on organizations, achieved largely through applied principal-agent theory, was originally conceived by economist Nobel Laureate Ronald Coase in the early 1930s, who argued that markets and hierarchies, heretofore examined as separate topics, were in effect substitutes for each other. The factors in a specific firm or division of a firm that made one alternative superior to another were often associated with the differing costs of coordination (the costs associated with transactions among individuals). This perspective has been applied most widely in recent years by Oliver Williamson and, even more recently, by Walter Powell and others who extended the argument to include a third substitutable form of organization, namely alliances or networks. Although the nature of the research carried out by these economists was often less quantitatively grounded, it nonetheless has had a profound effect on managerial thinking in all organizations, including thinking about education. They provide many of the arguments about current educational policies involving outsourcing, public-private partnerships, charter schools, vouchers, and accountability systems.

Regulation through standards- and outcomes-based accountability has become, in rhetoric at least, the prevailing mechanism favored by policymakers, although it takes different forms in different states (see Figlio & Ladd, this volume). For example, most accountability systems explicitly measure student outcomes, and they make information on outcomes available to parents and the public in order to permit judgments about school performance. In some cases, states allow parents some ability to respond to this information (e.g., through the receipt of supplemental services or through switching schools). Some schemes incorporate explicit monetary rewards for school leaders and/or teachers based on student performance.

The passage of the No Child Left Behind Act of 2001 (NCLB) brought national attention to the issue of accountability in schools and in the process drew much greater attention to the incentives influencing the behaviors of educators, especially teachers, principals, and school district superintendents and boards. As such, NCLB accountability embodies much of the framework that economists refer to as "agency theory" (Ferris & Winkler, 1991; Moe, 1984). Per agency theory, principals (superiors in organizations, e.g., school superintendents) seek to ensure that agents (subordinates in organizations, e.g., school principals) carry out the principal’s goals, in recognition of four problems that often arise to varying degrees in the vast majority of relationships between superiors and subordinates. An *adverse selection problem* occurs when principals (e.g., school superintendents) are not fully informed about the agents’ (e.g., school principals’) abilities and values and select agents that are not the best candidates. A *diverse objectives problem* occurs when agents pursue their own objectives at the expense of the principals’ objectives. This problem is compounded when compliance is achieved only by costly monitoring and controlling of the agents. An *information asymmetry problem* occurs when information within the accountability relationship is not evenly distributed. The agent typically has the information advantage. Finally, a *weak incentives problem* occurs when principals lack sufficient decision rights to cause the agents either to share principals’ values or to behave as if they did. Many critiques of NCLB and state accountability implementation address, often implicitly, these classic principal-agent problems and argue for changes that address them (Hentschke & Wohlstetter, 2004).

Accountability systems vary along many dimensions, but their overriding purpose (effective delegation of
authority and responsibility) is generally accepted. While much of the initial debate over NCLB centered on whether standardized testing is the right measure of student performance, some evidence suggests that accountability policies have led to increases in student achievement, but the results are not always consistent across subjects and grades (Carnoy & Loeb, 2002; Chiang, 2009; Dee & Jacob, 2011; Figlio & Loeb, 2011; Figlio & Rouse, 2006; Hanushek & Raymond, 2005; Ladd & Lauen, 2010). See Figlio and Ladd (this volume) for a comprehensive review of the literature.

**Concluding Thoughts**

In this chapter, we argued that economics, as a framework for understanding the allocation of resources, and human and organizational behavior more generally, have made important contributions to the study of educational policy. The core importance of economics in education, especially vis-à-vis other social sciences, lies in its conceptual structure and in the scientific approaches applied by economists to their research, fueled by the increasing availability of quantitative data on all aspects of the educational enterprise. As policymakers continue to focus on reform strategies that rely at least to some extent on economic foundations, and as information technology and the global economy continue to evolve in a way that appears to dramatically change the kinds of skills students need and the jobs available to them, the linkage between education and economics is likely to continue to grow.

As our review of some of the major education economics topics has suggested, the contributions and the impact of economics on education (past and projected) is uneven. The concept of human capital, for example, is well established, and has expanded beyond simple rate-of-return calculations to considerations useful for policymakers—for example, in determining the relative contributions of different types of institutions or programs of study. The interplay between markets, regulation, and individual and organizational performance in education has attracted considerable academic interest, and that is expected to continue, albeit frustrated by the absence of large-scale demonstrations of different structures of the sort that would permit more definitive statements about what works. Applying economics to policies affecting the design of education organizations has been a mixed bag. With some exceptions (e.g., on class size), production function and cost effectiveness studies have had limited policy impact. Agency theory and transaction-cost economics, intuitively appealing concepts, have not yet been rigorously applied to education organizations. However, as the education world begins to “flatten” (more privatized outsourcing, joint venturing, unbundling of education service provision, along with increased entrepreneurial behavior in education responding to greater consumer demand), this field should be expected to grow.

Needless to say, our view is that the increasing use of economics concepts in analyses of education policy is a positive development, in terms of helping to improve the overall rigor of education research, in providing educators and policymakers evidence on different strategies for improving schooling, and, indeed, with providing greater understanding of the effects of those strategies. As we look to the future, considering simultaneously the discipline of economics and the field of education policy, there seem to us to be a number of particularly promising areas where progress could be made.

First, the increasing issue of property rights and the economics of intellectual property will expand as for-profit providers enter K–12 education in greater numbers. Second, as income inequality continues to expand, the relationship of educational programs to other human and social services and poverty-reduction strategies will likely resurface as a research and policy area. Third, economic globalization will undoubtedly spread further into the education sector, as schooling at all levels increasingly is framed as a tradable, commoditized service, implying that aspects of macroeconomics might become more useful. This, in turn, suggests something more than the fact that education policymakers and economists will be increasingly approaching each other. Rather it suggests that both may well focus more attention simultaneously on education’s connections to the larger socioeconomic context within which it operates and on the incentives influencing the behavior of individuals deep within the education enterprise. It should be a stimulating period of intellectual activity.
Notes

1. See www.aefpweb.org and www.appam.org for sample agendas. AEFP provides for perhaps the biggest gathering of education economists. The AEFP journal Education Finance and Policy has been edited to date by three editorial teams, all economists of education. The largest educational research organization, American Education Research Association (AERA) (www.aera.net), remains relatively immune to the influence of economists, although its leading applied journal Educational Evaluation and Policy Analysis occasionally features research authored by economists and has economists serving on its editorial board.

2. Dozens of studies by economists use these data. For example, the data were used to examine issues of teacher quality and certification (Ehrenberg & Brewer, 1994; Goldhaber & Brewer, 1997, 2000); tracking (Rees et al., 1996); returns to education (Brewer et al., 1999; Grogger & Eide, 1995; Eide, 1994, 1997); the role of community colleges (Hilmer, 1998); the determinants of student achievement (Eide & Showalter, 1998); teacher race, gender, and ethnicity (Ehrenberg & Brewer, 1994); and numerous other topics.

3. The search for methods to approximate experimental designs has been advanced greatly in recent years. For example, regression discontinuity design (RDD), borrowed from psychology, has been used in a number of recent educational policy studies by economists (Jacob & Lefgren, 2004a and 2004b; Chay et al., 2005). A recent book by education economists Murnane and Willett (2011) addresses multiple methods for obtaining causal estimates in education research, including randomized experiments and natural experiments.

4. The screening hypothesis, for example, posits that schooling is just a mechanism through which individuals with traits that are valued by employers are sorted into jobs that pay higher wages. In this model, more schooling may just be a way to demonstrate to employers that workers are disciplined and motivated, as opposed to representing increased human capital attainment.

5. Friedman’s impact on education policy debates was fueled at least as much by the manner in which the ideas were communicated as by the idea itself. Despite Friedman’s contributions as an author, arguably most people became acquainted with his ideas through television, especially his 1980 10-part series, Free to Choose, which aired three times on public television and continues to air today via free Internet video-stream (www.ideachannel.tv). (See John Fund, “TV’s Evangelist for Capitalism,” Wall Street Journal, 1/31/07, p. D10.) The ongoing policy debate over education vouchers illustrates the complexity of linkages between research and policy.

References


Gill, B. P., Timpane, P. M., Ross, K. E., & Brewer, D. J. (2001). Rhetoric vs. reality: What we know and what we need to know about vouchers and charter schools. Santa Monica, CA: RAND.


The Past, Present, and Possible Futures of Educational Finance Reform Litigation

WILLIAM S. KOSKI AND JESSE HAHNEL

For some 40 years, the courts have played a significant role in shaping educational finance policy. Seizing upon arcane and often indeterminate state constitutional language, state supreme courts have invalidated the educational finance schemes of state legislatures and ordered those systems reformed in accordance with constitutional strictures. Through 2012, school finance lawsuits had been filed in 45 of the 50 states, with challengers prevailing in 27 of 47 cases that resulted in a judicial decision (National Access Network, 2012a, 2012b). Although early litigation focused on the development of the right to equal per-pupil funding, or at least a school finance scheme not dependent upon local property wealth, more recent litigation has sought to define qualitatively the substantive education to which children are constitutionally entitled.

This chapter explores the intellectual and legal foundations of the school finance reform litigation of recent decades, the legal and reform theories that have evolved in those litigations, and the potential directions in which this litigation might move in the future. Though the future of school finance reform through the courts remains uncertain, there can be no doubt that this experiment in judicial federalism has generated heated debate and affected education finance policymaking in those states in which courts have intervened.
Rooted in the hallowed principle of local control, public schools have traditionally relied on local property taxes as their primary funding source. Naturally, as property tax bases vary among districts, so do the property tax revenues available to schools. Although the extent of the interdistrict inequality in educational funding resulting from this property-tax-based system had long been recognized by educational finance experts, it received little outside attention until the late 1960s, when social activists and scholars began to notice the differences in educational resources available to students in different districts. Buoyed by victories in the Warren Supreme Court, these reform-minded scholars and activists began to look to federal law and litigation as potential tools to remedy the unfair distribution of school funding.

The Legal Environment

Prior to the landmark *Brown v. Board of Education* decision in 1954, courts rarely intervened in educational policymaking and practice. With that decision, the U.S. Supreme Court ushered in the Equal Protection Revolution that quickly spread from the unlawful segregation of children based on race to the differential treatment of other “suspect classes,” such as women and religious minorities, and the state’s denial of other “fundamental rights,” such as the right to vote, to an attorney in criminal cases, and, perhaps, to equal educational opportunity.

By the time scholars and advocates began turning their attention to the issue of unequal school funding, the U.S. Supreme Court had begun employing two distinct approaches to claims asserted under the Equal Protection Clause. The first and more relaxed standard of review under the Equal Protection Clause, known as the “rational review” test, upholds legislation so long as it reflects some rational relation between the state’s policy objective and the means the regulation uses to achieve that objective. Most legislation falls under this category. The second approach, requiring closer scrutiny of the law by the Court, is triggered when either a “fundamental right” is at stake or the state employs a “suspect classification.” Legislation subject to strict scrutiny is unconstitutional unless the state provides a compelling interest to which the challenged legislation is narrowly tailored as well as showing that the interest cannot be satisfied by any other means.

Whether education was a “fundamental right” was a question that had not been considered by the Court, but *Brown* had deemed it “perhaps the most important function of state and local government.” And while race was clearly recognized as a suspect class, by the late 1960s the Court had also begun to show a marked antipathy toward legislative classifications that discriminated on the basis of wealth. The legal groundwork had been laid. To scholars and advocates alike, school finance systems that provided fewer educational opportunities to children solely because they lived in property-poor communities appeared easy targets for this new jurisprudence. However, unlike a poll tax, the courts could not simply strike down the school finance scheme without providing guidance towards a constitutional replacement. At least, this was the conventional wisdom on the subject.

Strategies and Proposed Standards

In the late 1960s, several legal scholars and advocates began preparing the assault on school finance systems that provided vastly different educational opportunities to children. Although differing slightly on the details, all agreed that the legal basis for the attack was the Equal Protection Clause of the Fourteenth Amendment, while the proper forum would be the federal courts, as those courts seemed more willing to protect rights and liberties than their relatively complacent state counterparts. Where these scholars and advocates disagreed was in their interpretation of the specific constitutional wrong in the system and the judicial standards for constitutional compliance—i.e., the meaning of “equality of educational opportunity.” From this early thinking,
four contenders emerged: per-pupil spending equality or "horizontal equity," needs-based funding equality or "vertical equity," equal opportunity for an equal outcome or "effective equality," and the "fiscal neutrality" principle. This intellectual history is reviewed here because the questions raised by these early thinkers are still debated today.

One Scholar, One Dollar—Horizontal Equity. As a doctoral student, Arthur Wise became one of the early architects of the assault on the inequality produced by educational finance systems (Wise, 1967). To Wise, the central evil of educational finance schemes was their classification of students based upon the accident of geography and socioeconomic status. He reasoned that because education finance schemes classify students on the basis of the school district in which they reside, and because such classification largely determines the quality of the educational opportunity students receive, educational finance schemes that rely on local property tax bases unlawfully discriminate against children in low-property-wealth districts.

Mindful that courts would have to fashion a definition of equal educational opportunity to guide legislative remedies, he reasoned that they would most likely select a "negative definition" of equality of educational opportunity. Such a negative definition would require that a child’s educational opportunity should depend upon neither her parents’ economic circumstances nor her location within the state. The difficulty with this definition, Wise noted, is that it provides legislatures little guidance as to what constitutes a constitutional educational finance system. To be safe, Wise concluded, courts would likely adopt the "basic standard of equal dollars per pupil" (p. 159). Simplicity of application aside, however, the "one scholar, one dollar" standard appeared to many, including Wise, unsatisfying. It failed to account for the differential costs of doing business among districts, the differing needs of students, and the differing pressures on municipal budgets for social services. To rectify this deficiency, Wise suggested that courts might stray from this absolute equality standard to allow deviations in spending for different classifications of students.

Student Needs—Vertical Equity. Writing at about the same time as Wise, Harold Horowitz of the University of California, Los Angeles Law School was crafting a slightly different legal theory to attack educational finance schemes and preparing an arguably more ambitious standard for equality of educational opportunity under the Fourteenth Amendment (Horowitz, 1966; Horowitz & Neitring, 1968). Horowitz argued that equal protection jurisprudence could support a claim to strike down the state’s educational finance scheme where “a school board, though providing substantially the same educational programs and services in all schools, fails to provide programs and services which adequately compensate for the inadequate educational preparation of culturally deprived children” (Horowitz, 1966, p. 1148). Relying on empirical evidence that children in schools in “disadvantaged” neighborhoods perform poorly on academic achievement tests and receive fewer educational resources, Horowitz maintained that such children could enjoy “equality” only if they received “special programs, adapted to the specific needs of these children” (pp. 1166–1167). As a judicial standard, however, vertical equity was thought to be unmanageable, as it seems to require a student-by-student analysis and remedy.

Equal Opportunity for an Equal Outcome—Effective Equity. Perhaps the most aggressive standard for equality of educational opportunity to arise from the early Equal Protection scholarship is David Kirp’s (1968) call for effective equality. Kirp argued that “[a] reconsideration of effective equality in the light of recent and extensive educational research studies ... suggests that the state’s obligation to provide an equal educational opportunity is satisfied only if each child, no matter what his social background, has an equal chance for an equal educational outcome, regardless of disparities of cost or effort that the state is obliged to make in order to overcome such differences” (p. 636). To achieve this goal, two remedial schemes appeared to Kirp most promising: integration and resource reallocation aimed at effective equalization. Kirp argued that redistricting local school districts such that poor and minority youth would be integrated among their wealthier and whiter peers would “do most to better the chances of the poor, presently locked into predominantly lower class schools” (p. 661). But what about those districts for which such redistricting would be politically or geographically infeasible due to the sheer density of concentrated poverty among minority children and the resistance of wealthy suburbs? Kirp’s response was reallocation of resources pursuant to the principle of effective equalization—resources should be allocated to ensure children of different social backgrounds have the same academic success.

Theoretically, a meaningful distinction exists between the needs-based standard proposed by Horowitz and the outcomes-oriented standard proposed by Kirp. Horowitz would have the state compensate for educational
deprivation and needs without regard to outcome, whereas Kirp’s model—much like modern adequacy “costing-out” models discussed below—would focus on outcomes and the resources necessary for each student to reach the same high outcome. In practice, however, the connection between specific educational inputs and outcomes was unknown, creating a great deal of ambiguity in judicial standards. This ambiguity of standards was and is inevitable where theory out-paces empirical knowledge of what it takes to provide equal chances for equal outcomes or an adequate education.

**Fiscal Neutrality.** Jack Coons, William Clune, and Stephen Sugarman saw fiscal neutrality as the remedy for ambiguity. Less than two pages into their seminal work, *Private Wealth and Public Education* (1970), Coons, Clune, and Sugarman set forth their modest and clear standard for what would constitute a constitutional provision of educational opportunities within a state: “The quality of public education may not be a function of wealth other than the wealth of the state as a whole” (p. 2). What they then called Proposition One, and what would later be dubbed the “fiscal neutrality” principle, is a simple negative statement of what the state could not do—discriminate against students on the basis of the wealth of the community in which they live. Mindful of the complexity and contradictions inherent in defining equality of educational opportunity, Coons, Clune, and Sugarman designed this principle in a way that boils down to one simple measure: dollars. The availability of those dollars could not depend upon the wealth of one’s neighbors. Because the fiscal neutrality principle prohibited, rather than demanded, certain forms of state action, it allowed the courts to spark a major reform in educational finance policy while permitting the legislature to tackle the intricate difficulties of designing a fair and efficient system. A court could at once be activist and restrained. Finally, the negative statement of fiscal neutrality largely sidestepped the complex and ever-controversial issue of whether and how money matters in education, then known as the cost-quality debate. Under the Coons, Clune, and Sugarman formula, there was no need to demonstrate the link between educational resources and educational outcomes.

The fiscal neutrality principle did not mandate compensation for prior inadequate schooling, “cultural disadvantage,” or natural (in)abilities. Nor did it prevent some localities from choosing to spend more on their children’s education than others, so long as that choice was not dependent upon the wealth of a municipality. Indeed, Coons, Clune, and Sugarman saw the fact that some communities could tax themselves at higher rates to provide more educational resources to their children as a strength of their proposal. It would encourage educational experimentation, enhance local control, and recognize the independence and liberty interests that communities and parents should enjoy.

Unfortunately, the fiscal neutrality principle could do very little for those districts that needed the most help. By the late 1960s, educational failure had become synonymous with large, urban, minority districts. Children in such districts often faced multiple handicapping conditions, ranging from deep and persistent poverty to racial and cultural isolation to greater rates of physical, emotional, and mental disabilities. Paradoxically, those districts often enjoyed greater than average commercial or industrial property wealth; the problem was not the tax base, but the tax rate. Urban residents already taxed themselves to the limit to pay for municipal services that included amplified law enforcement, social services programs, and waste disposal. Suffering from such “municipal overburden,” urban communities simply could not afford to increase their taxes.

Yet fiscal neutrality as a principle was unconcerned with this problem. Judicial modesty and manageability were the touchstones for judicial intervention and the guiding principles behind Proposition One. The courts should apply only a negative test for constitutionality of an educational finance system, and they should refuse to prescribe specific components of equality of educational opportunity. This decision was best left to the legislature, Coons, Clune, and Sugarman argued.

This modesty put Coons, Clune, and Sugarman directly at odds with the more ambitious proposals to equalize opportunities of rich and poor children. From the work of Wise, Horowitz, Kirp, and Coons, Clune, and Sugarman, four theoretically distinct principles for judicial intervention in educational financing emerged. This chapter now considers how the courts grappled, and continue to grapple, with these standards.
The Past (Part II): The Evolution of Educational Finance Reform Litigation

Presented with an issue ripe for reform and armed with coherent and potentially winning legal strategies, educational finance reform advocates took their cases to court. Since 1968, according to the standard narrative, school finance litigation has developed in three waves (Heise, 1995b; Thro, 1989). Though the three waves are hardly monolithic and may be criticized for their descriptive accuracy (Koski, 2003; Ryan & Saunders, 2004), this standard narrative provides a common language to consider the shifting legal underpinnings of school finance litigation. In addition to the shift in legal doctrine, the wave narrative traces a shift from equity to adequacy in the distributional paradigm underlying school finance reform. This section describes the evolution of educational finance reform litigation in three waves.


Launched in the late 1960s, successful school finance litigation initially adopted the strategies developed by Wise and Coons, Clune, and Sugarman by focusing on the federal constitution’s Equal Protection Clause and the theory that per-student funding should be substantially equal, or at least not dependent upon the wealth of the school district in which the student resided. After enjoying initial success in at least two federal district courts and the California Supreme Court in Serrano v. Priest (Serrano I), the federal equal protection theory was quashed by the U.S. Supreme Court in San Antonio Independent School District v. Rodriguez.

At issue in Rodriguez was Texas’s system of educational finance, which relied almost exclusively on local property tax wealth and resulted in local school districts receiving radically unequal levels of educational funding. The questions before the Court were whether such a system violated the federal Equal Protection Clause and, more specifically, whether poor children in poor school districts formed a suspect classification, or whether education was a fundamental interest under the federal Constitution. Finding neither a suspect classification in children who lived in property-poor school districts nor a fundamental interest in education, a 5–4 majority of the Court applied the “rational relationship” test to Texas’s school finance plan and held that the state’s interest in local control over education easily supported the school funding scheme, unequal as it was. Though the Court left open the door to a federal constitutional claim against a state policy that deprived children of some basic floor of educational opportunity, Rodriguez effectively shut the door on federal school finance litigation under the U.S. Constitution to date.

The Second Wave: State “Equity” Litigation (1973–89)

Undaunted and capitalizing on the federalist structure of the judicial system, school finance reformers turned to state constitutions as sources of educational rights and school finance reform. Only 13 days after the Supreme Court handed down Rodriguez, the New Jersey Supreme Court ushered in the second wave of school finance cases with its discovery of educational rights in state constitutions. Although the Robinson v. Cahill court based its decision solely on the state’s education article, which imposed on the state legislature a duty to provide a “thorough and efficient” education to the state’s children, the critical aspect of the case was the newfound reliance on state constitutional arguments. Thereafter, most state high courts relied heavily on their state education article, at times employing it in conjunction with the state’s constitutional equality provision, when finding the state’s school spending scheme unconstitutional.

The essence of the claim in second-wave cases was the equity of school funding schemes. Specifically, the courts primarily sought to achieve either horizontal equity among school districts, such that per-pupil revenues were roughly equalized by the state, or at least fiscal neutrality, such that the revenues available to a school district would not depend solely on the property wealth of the school district. Unfortunately for plaintiffs in second-wave cases, the courts were mostly unreceptive to their claims: plaintiffs prevailed in only seven of the 22 final decisions in second-wave cases.
Beyond the win-loss record, several modest conclusions can be made as to the impact of second-wave, equity-minded educational finance litigations. First, in those states in which the state’s high court overturned the educational finance system, per-student spending across districts has become more equal (Evans et al., 1997, 1999; Murray et al., 1998). Second, this greater equity has in part been realized by greater funds being targeted to less wealthy school districts (Evans et al., 1997, 1999; Murray et al., 1998). Third, while some have argued that this increased equity has come at the expense of limiting overall growth in educational spending or reducing the state’s educational spending compared to other states, others have concluded that educational spending in the wake of a successful challenge to the school finance scheme increased school funding. Finally, a declaration that the educational finance system is unconstitutional typically leads to greater centralization in educational spending.

The Third Wave: State “Adequacy” Litigation (1989–Present)

The third wave was launched by the Kentucky Supreme Court in 1989 when it found in the education article of its state constitution not an entitlement to educational equity, but rather an entitlement to a substantive level of educational quality. Interpreting its thorough and efficient clause, the court held that the state legislature must provide its students with an adequate education, defined as one that instills in its beneficiaries seven capabilities, including, for example, sufficient oral and written communication skills to enable them to function in a complex and rapidly changing society. Though equity litigation has not been abandoned, Rose is considered the bellwether for the legal and rhetorical shift from equity to adequacy (Thro, 1989).

“Adequacy” as a distributional principle differs from any of those proffered by the early school finance scholars seeking to define “equality of educational opportunity.” An adequate education is understood to mean a specific qualitative level of educational resources or, focusing on the outcomes object, a specific level of resources required to achieve certain educational outcomes based on external and fixed standards. It is a measure that does not compare the educational resources or outcomes of students with each other; rather, it looks only to some minimally required level of resources for all students. Notably, in the context of adequacy lawsuits, the very same education articles that supported equity claims in the second wave would now be deployed for adequacy claims in the third wave.

One might argue that the move from equity to adequacy was a strategic necessity. Rather, the adequacy principle based on state education articles possesses many advantages over the equity principle based either on state equality provisions or education articles. First, by relying upon the education provision of the state constitution, judges would be less likely to create spillover effects in other areas of public policy. Changing the black-letter law of equal protection might invalidate not only locally financed education, but all other locally funded government services—a decision the scope of which courts were unprepared to order.

Second, adequacy arguments seem to flow naturally from the language of education articles, which generally require that the legislature provide a "thorough and efficient," “uniform,” or even "high-quality" education to its children. The court need not bend the language of these provisions beyond recognition to reach the adequacy standard or search for elusive "fundamental rights" and "suspect classes."

Third, a standard that relies on absolute rather than relative levels of educational opportunity would, at least in theory, avoid the ire of the state’s political and economic elite. A constitutional floor of adequacy permits local districts to provide their children more than what the court deems an "adequate" education. Similarly, an adequacy standard seems to intrude less upon the value of local control. The decision-making authority of well-to-do districts need not be curtailed simply because of a court order to the state that a poor school district be provided resources. Indeed, giving that school district the financial wherewithal to improve itself enhances local control.

Fourth, an adequacy standard may, at first blush, simply be more appealing to certain norms of fairness and opportunity. Modern American society views education as a key to economic success and social mobility. It is not much of a stretch to say that social and economic inequality are better tolerated in this country because Americans believe that the necessary tools for success are provided through public education. When one learns that some children are not receiving even the minimally adequate education that will help them better their lot, one feels that an injustice has been perpetrated. But Americans do not seem to feel this way if one child—most
often their own—receives a better education than another child, so long as that "other child" is getting an "adequate" education.

Finally, at least upon initial examination, the adequacy standard appears to enjoy a clarity that equality of educational opportunity lacks. Nettlesome concerns about input versus outcome equity and vertical versus horizontal equity are avoided. All the legislature needs to do is define what constitutes an adequate education and provide districts with the resources and conditions necessary to deliver that level of education.

Normatively, however, the shift from equity to adequacy might be seen as troubling. Education as a private good possesses strong positional aspects, with one’s employment opportunities and socioeconomic status depending (in large part) not on one’s absolute level of academic achievement but on one’s place in the educational distribution. Policies and constitutional holdings that mandate higher achievement but tolerate or even exacerbate already existing educational disparities serve only to further disadvantage the educationally underserved (Koski & Reich, 2006).

More pragmatically, the adequacy standard may provide no more clarity than ineffable equity standards. State constitutions provide legislatures, and ultimately courts, virtually no guidance as to what constitutes an adequate education. There is no agreed-upon list of public education goals. (Is it producing civic-minded democratic citizens, or productive contributors to the economy?)

There is no standard for the skills, competencies, and knowledge necessary to serve those goals of an adequate education. Finally, even if the legislature and courts were to craft those standards from whole cloth, how do we determine what resources will produce the desired outcomes? And what background characteristics of students ought to be considered in distributing those resources (e.g., linguistic, economic, and/or genetic disadvantages)? This chapter next considers how courts have addressed these issues in modern adequacy litigation.

The Present: Issues in Modern Educational Finance Reform Litigation

This section explores modern "adequacy" school finance litigation with a focus on how that litigation has developed in conjunction with the standards-based reform movement. Concurrently, it examines the difficult question of how courts have approached the challenge of crafting a remedy for the constitutional deprivation of an adequate or equal education.

Adequacy Litigations, Standards-Based Reform, and the "New Accountability"

Parallel to recent adequacy litigation, state legislatures have embraced the now-inseparable policies of standards-based reform and accountability for student outcomes. Put simply, the standards-based reform movement has sought, among other things, to combat low educational expectations for poor and minority children. By establishing challenging educational content standards that define what all children should know and be able to do, standards-based reform aims to raise the level of all children’s achievement to what the state determines is "proficient" (read: "adequate").

Since the mid-1990s, standards-based reform or the push for accountability has been supplemented by an additional policy lever—accountability of schools and students for performance on standards-referenced achievement tests. This "new accountability" in public education provides for rewards or sanctions to schools, administrators, teachers, and students according to their success in meeting achievement goals. At one end of the spectrum, successful schools are provided with commendations and, sometimes, monetary rewards. At the other end, failing schools may be offered technical assistance and temporary improvement grants, while persistently failing schools may be subject to state takeover or reconstitution. At a minimum, school and district performances on standards-based assessments are published and subjected to public scrutiny.

Standards-based accountability programs, such as the federal No Child Left Behind Act, though promising to raise the performance of poor and minority children and close the achievement gap, are frequently criticized for failing to provide the necessary educational resources and conditions for all children to achieve at world-class levels (Elmore, 2003). This is where the new adequacy litigation and new accountability movements are
beginning to embrace each other in courtrooms. Scholars and advocates have argued that it is institutionally appropriate for courts to hold states accountable under state constitutional education articles for providing the resources necessary for children to learn at the levels authorized by legislatures and often established by executive branches.

Although no state court has gone so far as to constitutionalize state educational standards, many judges are citing as evidence of educational inadequacy the failure of students to reach proficiency on state-mandated tests. Whether at the point of identifying the substantive entitlement to an education (the skills and capacities all children should receive) or designing the appropriate remedy (costing-out an adequate education based on student need or providing specific interventions and programs geared toward achieving the standards-based outcomes), courts are beginning to compel policymakers to flesh out the substantive entitlement to an education, sometimes based on states’ own expected educational outcomes.

System-Wide Reform: What Is the Remedy for an Unconstitutional Funding Scheme?

Although some courts have held the issue of educational adequacy to be nonjusticiable, a matter best left to the legislative branch, these decisions are in the minority. As explored above, courts are increasingly finding it within their power, indeed their duty, to rule on the constitutionality of their state’s method of financing education. But often that is all they do. Having found the funding system in violation of their state’s constitution, separation of powers concerns have led most courts to simply instruct the legislature to fix the problem.

In some states, this interbranch dialogue has led to meaningful education reform. But where legislatures appear politically unmotivated or unable to enact a constitutional funding scheme, some courts have taken a more active role. Unwilling to be complicit in their state’s failure to provide a constitutionally adequate education, they have ordered a range of remedies, from simply imposing deadlines upon the legislature to ordering specific, comprehensive, system-wide reform. Some of these more expansive, and often controversial, remedies and their legal underpinnings are explored below.

Costing-Out. State funding for education has historically been a product of political deal making, economic pressures, and the struggle among competing interests for limited state resources. Costing-out is an attempt to tie educational funding to the actual amount needed to provide every child a constitutionally adequate education. Many legislatures have conducted costing-out studies independent of judicial intervention, but others have done so only in response to threats of litigation, settlements, judicial orders to enact a new and constitutional system of educational funding, or occasionally, specific orders to conduct such studies.

Costing-out remedies are a logical extension of adequacy claims. If the court finds the state to be violating its constitutional obligation to fund an adequate educational system, it must know how much more before it can order more funding. A costing-out study can inform that question. The pathbreaking case in this respect is Campbell v. State, in which the Wyoming Supreme Court directed the legislature to define the “basket” of education every Wyoming child should receive, to undertake a “cost of education” study to determine the actual cost of providing such a basket, and to fund such an educational system. Other state courts have followed, with New York, Arkansas, and Ohio all ordering the legislature to conduct costing-out studies as a prerequisite to funding an adequate educational system.

Although adequacy claims naturally lead to costing-out remedies, costing-out sometimes results from equity claims. In Arizona, a federal court twice ordered costing-out studies performed in response to Equal Education Opportunities Act (EEOA) claims, though these costing-out studies were done only for English language learners. Costing-out studies have also resulted from state equal protection claims, as legislatures act to provide a “rational basis” for their funding decisions.

As prevalent as costing-out studies have become, they are not without their critics, who focus not on the need for such studies but rather on the feasibility of obtaining accurate and reliable results. Nonetheless, recognizing the imperfections in the studies but also the need for a rational method of determining necessary educational resources, courts and legislatures continue ordering and implementing costing-out studies.

Programmatic Mandates. While many courts have hesitated to intrude on the legislative domain by directing them to fund education on the basis of a costing-out study, a few courts have gone further and directed their
legislature to fund and implement specific educational programs found necessary under the state constitution. The leading cases in this vein are New Jersey’s Abbott decisions where, after a period of “interbranch dialogue” spanning over 12 years, the court finally made clear (i.e., ordered) what was required. Among the most expansive of the programmatic mandates ordered by the court was implementation of Success for All, a whole-school reform program for elementary schools. Other mandated programs include interventions aimed at reducing dropout rates; school-to-work and college-transition programs; summer-school, after-school, and school nutrition programs for which there is demonstrated need; and art, music, and special education programs beyond those required as part of the reform plan. More recently, courts are going beyond K–12 education and mandating early enrollment kindergarten and preschool. This is hardly surprising: evidence suggests that remedial programs are more effective and less costly when intervention occurs at a young age (Barnett, Jung, & Lamy, 2005).

Remedies Focusing on Subclasses of Children. Just as some courts have ordered remedies focused on subsets of the educational offerings (programmatic remedies), others have ordered remedies focused on subsets of students. This most often takes the form of courts focusing adequacy opinions on the plight of at-risk children, with remedies tailored to “ensuring that ‘at-risk’ students are afforded a chance to take advantage of their constitutionally-guaranteed opportunity to obtain a sound basic [adequate] education.” In some instances, a case has been brought by students protected by specific legislative mandates. For instance, English language learners have prevailed under the EEOA and children with disabilities have prevailed under the Individuals with Disabilities Education Act (IDEA). The underlying claim—inadequate or unequal educational opportunities—remains the same, but the remedy—extra programs or funding—is directed toward a subclass of students.

Race is a subclass to which courts pay especially close attention, and when issues of educational equity and racial segregation mix the resulting remedy might be neither programmatic nor financial in nature. In a particularly innovative suit, plaintiff school-children in Hartford, Connecticut, claimed that the racial isolation of the Hartford public schools prevented them from receiving substantially equal educational opportunity as guaranteed by the state’s education and equal protection clauses. Reading these two clauses together, the Connecticut Supreme Court agreed. It held the state’s creation of school boundaries to be the most important factor contributing to the racial disparity between Hartford and the surrounding suburbs and ruled that “the existence of extreme racial and ethnic isolation in the public school system deprives schoolchildren of a substantially equal educational opportunity and requires the state to take further remedial measures.” Underlying this ruling was an understanding that a child’s educational experience is defined by more than just the programs provided by the state; it includes, perhaps most importantly, the child’s interaction with his peers.

But the Connecticut case is an outlier, and even it did not force the state to redraw or redefine district boundaries. The parties agreed to a settlement requiring the state to create eight new interdistrict magnet schools in Hartford, provide additional seats in suburban schools for minority public school students from Hartford, and provide increased funding for interdistrict cooperative programs, a far cry from a truly interdistrict, regional solution. Moreover, in a similar suit, the New York Court of Appeals held the state not “responsible for the demographic makeup of every school district ... [as this] would ... subvert the important role of local control and participation in education.” It concluded that "if the State truly puts adequate resources into the classroom, it satisfies its constitutional promise under the Education Article, even though student performance remains substandard." This opinion envisions a very different future for judicial involvement in educational finance litigation, one in which the state is found to be supplying a constitutional educational system despite perpetual substandard performance—low performance being a product of socioeconomic factors for which children, not the state, are responsible.

Whether it be costing-out, specific programmatic remedies, or a focus on subclasses of children, courts and advocates appear to appreciate that increased funding for schools is insufficient. Inequities persist, and disadvantaged children remain without their constitutionally guaranteed adequate education. The following section explores the possible futures of educational finance litigation by considering suits brought on behalf of educational units other than the district—i.e., suits brought by individual schools, students, and states. The section concludes with thoughts on the continuing role, if any, of courts in educational finance reform.
The Possible Futures of Educational Finance Reform Litigation

As education litigation matures, its scope is expanding beyond simple challenges to statewide school funding schemes. Gone are the days when the only legal relationship in question is that between the state and its districts. Gone too are the days where educational resource and educational equity litigation focuses only on school finance policy, ignoring other policies that may exacerbate inequality. This section examines current and possible educational policy issues that have been or may be raised in matters nominally deemed educational finance lawsuits.

The School as the Focal Point for Litigation and Remedial Efforts

If a state has a constitutional obligation to provide every child an adequate education, why should an individual school not bring a suit claiming inadequate resources and funding, or an inexperienced and ineffective teaching staff? Manageability concerns have led some courts to reject this legal approach on the basis that, absent evidence that the state is underfunding the district as a whole, the school has no claim against the state. But what if the state is not underfunding the district? What if the district is misallocating its resources and underfunding the school? Or what if district or state policy results in an unequal distribution of resources among schools within a district? Can children in those schools bring a lawsuit?

In certain instances, it may be the case that state and district policies result in an unequal distribution of resources at the school-site level. Take the case of Reed v. California brought on behalf of children in certain high-poverty schools in the Los Angeles Unified School District (LAUSD) who suffered the direct and serious consequences of a rigid teacher employment law. As a result of California’s budget cuts during the 2008–2009 academic year, thousands of teachers in the LAUSD received pink slips, letting them know that they were being laid off. Because these teachers were at the bottom of the seniority ladder and because California state law prescribed a last-in, first-out layoff process, those layoffs disproportionately affected low-performing schools that had higher concentrations of junior teachers. The Reed complaint focused on three middle schools in which one of the schools lost 70 percent of its faculty, while the other two lost more than 40 percent. Some schools lost no teachers. At the beginning of the 2009–10 school year, many students in those three middle schools entered classrooms with substitute teachers who would be there for months. Advocates on behalf of children in those schools sued LAUSD and the state, alleging that the seniority-based layoff rules denied children access to an equal education. The trial court agreed and found that the reduction-in-force policy violated children’s equal protection rights in those schools. Although the case is currently unresolved due to the appellate court’s rejection of a settlement agreement, the Reed litigation demonstrates the vulnerability of policies that create inequity among schools within districts.

More generally, intradistrict funding inequality may be susceptible to legal action. Research suggests that "variations of per pupil funding within districts are often greater than the within-state variations that have been found unconstitutional" (Roza & Hill, 2004). A primary reason for such intradistrict inequality is the practice by which teachers are paid "directly" by the district, with each school given the authority to hire a number of teachers proportionate to the school’s enrollment, independent of salary constraints. Because teacher’s salaries are also usually independent of the difficulty of their teaching assignment, there is an incentive among teachers to teach in easier, more affluent schools. Moreover, the more “affluent” schools are often staffed with more experienced, more expensive teachers. Thus, schools serving low-income, minority students frequently receive substantially less funds than schools serving more affluent, Anglo students, even after Title I and other categorical funds are taken into account. While suits challenging the constitutionality of funding inequality are rare, more might be forthcoming.

Interstate Inequality as the Focal Point for Litigation and Remediation

Even when adjusted for cost-of-living differences, per-pupil spending varies dramatically between states, with states containing more economically disadvantaged, minority, and English language learner students likely to
spend less (Liu, 2006). In fact, “disparities between states accounts for more of the variation in district per-pupil spending nationally than disparities within states” (Liu, 2006). While this could reflect a preference for less educational spending, data indicate that it is more a function of ability than willingness to pay (Liu, 2006). Absent substantial federal aid aimed at equalizing educational opportunities, children in wealthier states will continue receiving educational opportunities superior to those in poorer states.

The difficulty is that the likeliest source of legal rights to remedy interstate inequality, the U.S. Constitution, appears to most advocates to have been foreclosed by the Rodriguez decision. However, then Berkeley law professor and current California Supreme Court Justice, Goodwin Liu, suggested a legal argument for establishing a Congressional “duty”—though not an actionable “right”—to ameliorate such interstate inequalities. Arguing the Citizenship Clause of the 14th Amendment constitutes a “font of substantive rights,” with the affirmative nature of the Clause acting to expand Congress’s enforcement power beyond protecting national citizenship from state invasion, he concludes that “the constitutional grant of congressional power to enforce the national citizenship guarantee implies a constitutional duty of enforcement” (Liu, 2006). While Liu uses legislative history from the period directly subsequent to enactment of the 14th Amendment to buttress the validity of this interpretation (Liu, 2006), recent increased involvement of the federal government (through legislation such as the No Child Left Behind Act and the Individuals with Disabilities Education Act) in the traditionally local matter of public schooling makes claims of adequate educational opportunities being a component of citizenship all the more reasonable. While not providing a source of action to aggrieved individuals, the paper serves to remind national legislators of their moral and constitutional responsibility to abet interstate educational equity.

“Efficiency” Claims in Educational Finance Litigation

Dovetailing with the charter school movement has been a growing call to deregulate our traditional public schools by providing local administrators with greater discretion over educational spending (e.g., elimination of rigid categorical funding streams in favor of “block grants”) and greater flexibility in the hiring, firing, compensation, and assignment of teachers (e.g., relaxation of teacher due process protection, seniority assignment rules, and the step-and-column salary schedule). While much of this activity has occurred in state legislatures, the call for deregulation, sometimes dubbed “efficiency,” has begun to enter the courtroom. The Texas school finance litigation and California’s Vergara litigation are exemplary.

In the long-running Texas school finance litigation saga, a group of some 500 school districts sued the state in the wake of its draconian budget cuts for the 2011–12 school year, alleging that those cuts violated the state’s clearly established obligation to provide children an adequate education. Shortly thereafter, a group of students, parents, and taxpayers, calling themselves the “efficiency interveners,” joined the lawsuit to seek, as part of any remedy, changes to Texas law that would allow administrators and charter schools greater flexibility, including removal of the cap on the number of permitted charter schools, greater accountability for how dollars are spent in schools, and increased flexibility to fire poor performing teachers. The theoretical basis for such intervention may signify either a departure from or maturation of the central principle behind adequacy litigation: additional resources are necessary to improve results. The interveners seemed to claim that additional resources may or may not be necessary, but they certainly will not be effective without sufficient flexibility in and accountability for how the resources are deployed. Although the trial court declared the funding cuts unconstitutional and rejected the interveners’ claim, the Texas litigation was a first shot across the bow for those seeking “efficiency” in school spending.

Perhaps the logical next step for efficiency advocates is California’s Vergara litigation, which doesn’t seek additional educational resources at all, but rather contends that certain teacher employment laws have conspired to deny certain California children access to an equitable education. In that case, a group of students and parents—funded by a wealthy Silicon Valley entrepreneur and supported by former Washington, DC, schools chancellor Michelle Rhee’s organization, Students First—have specifically alleged that California’s teacher tenure laws, due process protections, and seniority-based layoff rules have conspired to deny them equal protection of the laws. The remedy they seek is not additional resources for those affected children, but rather the elimination of those teacher employment laws. Although the remedy differs greatly from traditional school finance cases, the legal theories are much the same. While that case’s outcome remains uncertain, it
seems likely that future educational resource litigation will not be focused solely on calls for “more money” but will additionally consider how that money is best spent to improve student outcomes.

The Continuing Role of the Judiciary in Educational Finance Policymaking

Though advocating a litigation strategy to attack the manifest unfairness of property-tax-based educational finance schemes, Coons, Clune, and Sugarman (1970) were equally wary of the courts intervening in the complex social policy arena of educational finance without clear standards. As the previous section demonstrates, courts have frequently failed to heed that advice and have become entangled in often lengthy and often contentious policymaking “dialogues” with state legislatures. The inevitable question is whether, and if so how, courts will help shape the future of educational finance and policy.

There is no doubt that the role of the judiciary in what Abram Chayes (1976) famously called “public law litigation” differs dramatically from traditional litigation among private parties. Ever since at least Brown v. Board, courts, under the authority of state and federal constitutional and sometimes statutory mandates, have been called upon to reform and superintend complex institutions such as schools and prisons in equally complex social policy arenas such as child welfare and police practices. In educational finance reform cases, state supreme courts have become deeply enmeshed in the policymaking process. Whether it is exercising the judicial “veto” over a school finance scheme, directing that a legislature “cost-out” an adequate education, or even prescribing specific educational reforms such as preschool for all, courts are making forays into policymaking that was traditionally left to legislatures.

For some, this dialogue between the judiciary and the political branches is a logical, if not healthy, role for the branch charged with protecting constitutional values (Jaffe, 1991). James Liebman and Charles Sabel (2003) have recently argued that in these school reform cases the judiciary is beginning to create public forums in which the political branches, educational insiders, and “new publics” (coalitions of civic-minded outside reformers) can “discuss comprehensive reforms of American education that draw on linked innovations in school governance, performance measurement, and the reconceptualization of the teaching profession and pedagogy” (p. 280). In this model, courts assume a coordinating and oversight role, enabling the new publics to reform educational finance policy once powerful interests have been “disentrenched” by judicial decisions. Whether this will become the norm in educational finance litigation is, however, unknown, as some are much less sanguine about this new role for the judiciary.

Indeed, the courts have come under increasing scrutiny and even attack for intervening in and invalidating educational finance policy (Starr, 2005). Courts have been criticized for crafting unmanageable constitutional standards, straying beyond their expertise, ignoring the separation of powers, and calling into question the very legitimacy of the judicial institution. Faced with such challenges, courts may choose a very different future from that lauded by Liebman and Sabel. Since the inception of educational finance reform litigation, many courts have avoided the complex waters of school finance policy by invoking the “political question doctrine” or citing concerns about the separation of powers. Perhaps a more chilling example for would-be judicial reformers, however, is the risk of being ineffectual in the reform process. Some argue that the Ohio educational finance reform litigation provides a cautionary tale, a tale in which, after repeatedly striking down the state’s educational finance reform system and being rebuffed by the legislature (and potentially facing a constitutional crisis), the Ohio Supreme Court relinquished jurisdiction over the matter.

Whether the judiciary will ambitiously pursue the role of participant in a school reform dialogue with policymakers and other “new publics” or whether courts, feeling fatigued from 30 years of litigation and not infrequent legislative recalcitrance, will withdraw from the school finance debate remains to be seen. What is known, however, is that the last few decades of educational finance reform litigation have permanently injected the “constitutional” values of equality and adequacy into that debate.

Notes


See, e.g., Ill. Const. art. X, § 1 (“The State shall provide for an efficient system of high quality public educational institutions and services.”).

See, e.g., N.J. Const. art. VIII, § 4, cl. 1 (“The Legislature shall provide for maintenance and support of a thorough and efficient system of free public schools….”).

See, e.g., Wis. Const. art. X, § 3 (“The legislature shall provide by law for the establishment of district schools, which shall be as nearly uniform as practicable.”).

See, e.g., Ill. Const. art. X, § 1 (“The State shall provide for an efficient system of high quality public educational institutions and services.”).

For a discussion of the difficulties in establishing an “adequacy” standard, see Koski and Levin, 2000.

In their seminal work on the measurement of equality in educational finance, Robert Berne and Leanna Stiefel later called this standard “horizontal equity” (Berne & Stiefel, 1984).

In Berne and Stiefel’s (1984) terminology, this is deemed “vertical equity.”

California could provide no compelling state interest for the local property-tax-based finance system nor demonstrate that the system was narrowly tailored to achieve the state’s interests. Although the court found the funding system unconstitutional under the federal Constitution, the court would later reconsider the matter and again find the state’s funding scheme unconstitutional under the state constitution. Serrano v. Priest, 557 P.2d 929 (Cal. 1976) (Serrano II).

This usually meant greater state-level involvement in educational funding through state-guaranteed tax base plans or, on rare occasion, state-backed equal yield plans, a.k.a. district power equalization, that sought to recapture “excess” revenues from wealthy districts.

See Evans et al., 1999, pp. 74–75 (noting that California has achieved finance equity through leveling down high-revenue districts); Joondeph, 1995 (concluding that California’s Serrano decision depressed educational spending in the state); Heise, 1995a (finding a negative relationship between judicial intervention in Connecticut’s school finance policy and overall state educational spending); Sonstelie et al., 2000 (“[S]pending per pupil in California between 1969 and 1998 fell about 15% relative to the average for the other states.”). It should be noted, however, that some of the evidence for this proposition comes from California—a state in which school funding has been further stymied by the property-tax-capping effects of Proposition 13. See Silva and Sonstelie, 1995.

See Hickrod et al., 1992; Evans et al., 1999.

Serrano I is widely recognized as the case all equity litigations sought to emulate. There, the California Supreme Court considered the now-infamous discrepancy in funding between the Baldwin Park and Beverly Hills school districts. In 1968–69, Beverly Hills enjoyed a per-pupil assessed valuation of $50,885, while the largely minority Baldwin Park suffered a $3,706 valuation. These disparities were naturally reflected in per-pupil expenditures: where Beverly Hills lavished $1,231.72 on each of its students, whereas Baldwin Park could afford to spend only $577.49 per student. This difference prevailed in spite of the fact that Baldwin Park taxed itself more aggressively than Beverly Hills. Based on the federal Equal Protection Clause, the California Supreme Court found that education was a "fundamental right" and poverty a "suspect classification." Therefore, judicial "strict scrutiny" should apply.

"Lines drawn on the basis of wealth or property, like those of race ..., are traditionally disfavored.

The state had not provided the constitutionally required "sound, basic education"); Montoy v. Kansas, 102 P.3d 1160, 1164 (Kan.

Florida and Rhode Island are typical of states holding any judicial ruling on educational adequacy to constitute an impermissible foray into the legislative realm. See, e.g., Coalition for Adequacy and Fairness in School Funding, Inc. v. Chiles, 680 So. 2d 400 (Fla. 1996); City of Pawtucket v. Sundlun, 662 A.2d 40 (R.I. 1995).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).

While most scholars view costing-out as an inexact and evolving methodology, some argue that it’s so susceptible, and subject to, manipulation, that it ought not be used by judicial decision makers (Hanushek, 2006). For example, the Professional Judgment Model produces different results depending on the professionals chosen, the Similar Schools approach is susceptible to data-based manipulation, and the Cost Function Model is highly sensitive to technical assumptions. Competing studies using similar methodologies produce radically different results (Guthrie & Springer, 2007).
identifying specific schools that do not meet minimum standards, plaintiffs do not allege any district wide failure … in seeking to require the State to assess and rectify the failings of individual schools, plaintiffs’ theory would subvert the important role of local control and participation in education”.


45. In a study of Baltimore, Cincinnati, and Seattle, high-poverty, low-performing schools employed teachers whose salaries were lower than average (Roza & Hill, 2004). In Houston, high-poverty, low-performing schools were found to be receiving significantly less funding than low-poverty, high-performing schools (Atiles & Roza, 2002).

46. To the authors’ knowledge only one such claim has been litigated: Rodriguez v. Los Angeles Unified School District, No. C 611–358, Superior Court of the State of California for the County of Los Angeles, with a consent decree entered August 25th, 1992.

47. See, e.g., Hornbeck v. Somerset County Bd. of Educ., 458 A.2d 758, 790 (M.D. 1983) ("[I]t is not within the power or province of members of the Judiciary to advance their own personal wishes or to implement their own personal notions of fairness under the guise of constitutional interpretation. The quantity and quality of educational opportunities to be made available to the State's public school children is a determination committed to the legislature..."); Britt v. North Carolina Bd. of Educ., 357 S.E.2d 432, 437 (Cl. App.), appeal dismissed 361 S.E.2d 71 (N.C. 1987) (holding that good law or bad law, wise or unwise, the question of what type of education to provide to North Carolinians is for the legislature, not the courts); Committee for Educ. Rights v. Edgar, 174 Ill. 2d 1, 28 (1996) ("What constitutes a 'high quality' education, and how it may best be provided, cannot be ascertained by any judicially discoverable or manageable standards"); City of Pawtucket v. Sundlun, 662 A.2d 40, 59 (R.I. 1995) ("The volume of litigation and the extent of judicial oversight provide a chilling example of the thicket that can entrap a court that takes on the duties of a Legislature.").


References


The Continually Evolving Political Context of Education Finance

JAMES W. GUTHRIE AND KENNETH K. WONG

Introduction

A popular perception of politics is that it is a process that determines who gets what, when or who pays, and who benefits. When viewed more abstractly, politics is a process by which a collective (a tribe, an organization, a community) allocates what it considers to be of value. The kinds of items or conditions that are valued include resources such as material objects, personal or public privilege, public symbols, access, power, and participation. All human societies have procedures for deciding how to allocate these and other resources. What distinguishes open and closed societies is the extent to which the political rules governing these decisions, and the breadth of political participation, are explicit and popularly agreed upon.

Although modern media and pundits periodically suggest that contemporary political activities are unseemly and political officials self-serving, an open society depends crucially upon a free flowing debate regarding allocation of resources. The alternatives to an open political system are dictatorships and oligarchies in which the same distributive dimensions exist but where the decision processes are dominated by far fewer and almost assuredly less civic-minded powerbrokers.

This chapter elaborates upon the political dynamics inherent in open or democratic societies and describes the operating dynamics that influence the manner in which education is financed.

Politics and Political Systems

Education in both K–12 and postsecondary sectors in the United States is an enormously expensive undertaking that accounts annually for 6 to 8 percent of the dollar value of the nation’s total annual production of goods and services, its gross domestic product. Large numbers of individuals are directly and indirectly affected by the education system; they range from parents, school and college enrollees, employees, and vendors to those more remotely engaged, such as citizens who pay taxes and vote on education issues. Given the magnitude of the resources at stake, the extensive spectrum of participants and benefactors, and the widely perceived long-run significance of education to the overall society, it is no wonder that executive and legislative branch officials—as well as a seemingly infinite number of interest groups and advocates, at the federal, state and local levels—spend vast amounts of time lobbying over, and deliberating upon, education issues.

The political process takes place within a context of overarching cultural perspectives and societal values, specialized spheres of interests and expertise, and governmental arrangements. This context can shape the nature of the political process and its outcomes and is, of course, reciprocal. The outcomes of the political process also shape the larger society and governmental arrangements.

Today’s political actions frame the societal context for tomorrow’s issues. David C. Easton (1953) was the first to portray the political process and its reciprocal relationship with the larger society as a system composed of varying parts, each of which is capable of exerting influence upon the other. Most notably, Easton captured the important interactions between the larger society, the political processes, the centrality of government, and
the feedback loop that results when a set of political outputs influences a subsequent set of political demands. In other words, the political issues of today, and the formal political system’s responses to them, set the stage for a subsequent wave, possibly a reactive one, of political dynamics.

Moreover, to add yet another layer of complexity, political dynamics can also be shaped by the particular issue at hand. For example, proposals to expand the number of charter schools, expand public aid to religious schools, or increase local school district property taxes are likely to motivate different sets of political actors and to lead to different forms of political engagement. A variety of policy spheres and specialized interests can catalyze political processes and structural arrangements and fuse them in unique combinations.

Politics involves various processes, values, interest groups, spheres of policy perception, and governmental arrangements. Even though these components are continually evolving and reciprocally shaping and reshaping each other, this chapter consciously separates them in order to describe the fundamental elements of politics more clearly. The reader should understand, though, that this separation is artificial and that political reality involves a booming, buzzing cacophony of complexity. Below is a practical illustration of a condition of long historical standing and continued contemporary significance that illustrates the interactions between the components of political systems (Eberling, 1999).

Illustrating the Interaction of Structure and Process

In 1647, the Commonwealth of Massachusetts enacted Ye Olde Deluder Satan Act. This colonial-era legislation specified that each community would have a specialized governmental body overseeing schools, that local school boards would be composed of laypersons, and that they would have property taxing authority. The fact that property wealth was ill distributed geographically was known but was not objectively quantified as are such metrics today. Hence, the inequity was considered of little consequence in colonial times.

Today, vast differences exist in the amounts of assessed value per pupil across different local school districts, and these contribute to substantial differences in local school district spending. Many years later, the fact that Massachusetts, the rest of New England, and much of the remainder of the nation adopted local school districts as a structural feature triggered heated legislative debate and litigation over matters of education finance equalization. Uneven wealth distribution in turn contributes to otherwise unusual coalitions of low-wealth school districts, such as those in rural and urban areas that, while having little else in common, cooperate in order to influence legislative and judicial financial distribution outcomes.

Public Values and Political Processes

Figure 4.1 depicts a triangle within a rectangle. Each side of the inner triangle represents a core value historically crucial to American democracy: equality, efficiency, and liberty. These three values are deeply embedded in the public ideology and political ecology of the United States, and each is continually and collectively reinforced in the nation’s public symbols and political rhetoric. The outer sides of the rectangle represent forces at work in the larger society. Changes in these external forces may trigger disequilibrium in the political system and provide an opportunity for a macro political shift in policy direction.
The three public values—equality, efficiency, and liberty—are often represented as symbols, and public officials frequently refer to them as ideals to be maximized. Whereas the three values are held dear in the abstract, they are often in direct conflict in practice. To pursue equality to its absolute metaphysical limit is to substantially restrict liberty. To eschew equality and to pursue only choice or liberty runs the risk of creating wealth and social class extremes sufficient to jeopardize social and political stability. Finally, while efficiency may well be furthered by choice and liberty, unfettered pursuit of efficiency may jeopardize equality.

Liberty, which is the ability of individuals and groups to choose and to maximize personal preferences, is often seen as the higher goal. Unfettered choice in schooling, however, might jeopardize civic unity or social cohesion and could exacerbate material and social inequality. Equality, the restriction of differences for its own sake, can be empty. To be sure, a democracy must strive to achieve and maintain equality of opportunity. Few contend, however, that absolute material equality, with all of its likely trappings of drab sameness, is an end in itself. If everyone had the same clothes, cars, houses, and food, there might be equality, but tedium might well be the order of the day. Absolute equality, or at least the pretense of such, was the hallmark of the former Soviet Union. Its ultimate downfall was a function of an inept system of individual and collective performance incentives, distance between ideological aspirations for citizen equality and the material corruption of its leaders, and its inability to provide consumers with choice.

Similarly, efficiency, the pursuit of maximal output while striving to minimize inputs, usually is taken to be an instrumental or mediating goal. Efficiency for its own sake has little intrinsic value. However, the conservation of resources so as to have more of something, be it material or psychic, would seem to contribute to enhanced choice or liberty.

Thus each of the three values, or the practical expressions sought by their proponents, competes with the other two in a political system. For much of the post–World War II period the American political system concentrated upon the pursuit of equality. This was particularly true in education, starting with school racial desegregation following the U.S. Supreme Court’s landmark 1954 decision in *Brown v. Board of Education*. Subsequent federal enactments such as the 1965 Elementary and Secondary Education Act (ESEA), the 1976 Education for All Handicapped Children Act, gender equity requirements contained in Title VI of the 1965 Higher Education Act, and the 2001 extension of the ESEA known as the No Child Left Behind Act (NCLB), state court school finance equalization decisions, such as *Serrano v. Priest* in California and *Robinson v. Cahill* in New Jersey, and immigrant education decisions such as *Plyler v. Doe*, illustrate the half century of policy concern for educational equality.

While equality was ascendant as a policy consideration between World War II and the turn of the 21st century, concerns for liberty and efficiency were by no means eclipsed. For example, tax limitation movements, charter school and voucher proposals, and demands for greater market play in education have been on the policy agenda at all levels of government since the 1980s. Contemporary concern for efficiency and productivity in the education system is also evident in components of the NCLB, federal Teacher Incentive Fund legislation, and state pay-for-performance teacher pay plans.

---

**Equilibrium Theory**
Systems, whether biological, ecological, or political, seek balance or stability and strive continually to adjust to their external environments. This tendency is known as homeostasis. A major function of the political system is to maintain a balance between the preferences of proponents of the three core values and their practical expression in day-to-day activities and conditions. For example, an emphasis upon equality that results in a restriction of choice or an erosion of liberty may provoke a reaction. Similarly, too great a concern for equality of school spending may motivate some parents to seek private schooling for their children, while too great a concern for liberty may alienate those who perceive themselves disadvantaged by choice and may lead to pressure for a restoration of concern for equality.

Within the value triangle depicted in Figure 4.2 are multiple interests, each of which may be associated with a different perception of what constitutes a just balance of the three values. This figure displays the interest spheres of parents, students, and the polity, each of which may have a unique view of what is equitable, efficient, or liberating when it comes to education policy.

**Micro Politics**

The “iron triangle of politics” influences strongly what has come to be labeled incremental or micro politics—the ceaseless day-to-day actions and adjustments of local, state, and federal officials and interest groups to maintain the stability of the political system. The iron triangle involves three categories of political actors: (1) self-interested advocates for change or direct recipients of prospective benefits, (2) policymakers who concentrate in this specialized policy sphere, and (3) executive branch officials and bureau workers who administer the programs connected with these interests. When it comes to incremental politics, these three parties are the dominant actors.

![Figure 4.2 The Triangle of Values With Interests.](image)

To illustrate the iron triangle in operation, consider a set of local school districts that desire a change in the state’s school finance formula. However altruistic the banners under which they mobilize, proponents are also likely to be organizational, and possibly personal, beneficiaries of the changes they seek. These distribution formula issues are unlikely to attract the attention of the larger political system involving political parties, the governor, the speaker, or the full legislature. Rather, these micro issues are more likely to be settled within the smaller domain of the iron triangle. However, if the desired formula alternations have substantial cost consequences and could lead to major redistributions of existing resources among districts, then the issue may exceed the boundary of the iron triangle and come to the attention of the larger political system.

More generally, much of the political system’s routine activity is engaged with maintaining a homeostatic condition. Micro adjustments are continually being made to finance formulas, to higher education student
financial aid, or to institutional appropriation levels so as to maintain stability in the system. On selected occasions, however, a more dramatic change takes place. The issue may become partisan and provoke political parties to take positions regarding the issue. A governor, a mayor, the full state legislature or the full city council may get engaged. Kingdon (1995) labels this latter condition "high politics."

**High Politics**

Certain dynamics can upset the value balance and shift the emphasis of a political system in such a way that it produces new or significantly altered policies. Returning to Figure 4.2, one can see a peripheral rectangle, each side of which represents one of four change vectors: technology, economics, demographics, and ecology. A significant change in any one of these forces can distort the public value balance, trigger a change in policies, and eventually provoke a search for a recalibrated policy equilibrium. Following are illustrations of this process at work.

Technological innovation is a major source of system imbalance and, hence, a principal trigger of new policies. Here is a dramatic example that had major consequences for the nation’s education system through the latter half of the 20th century. In the 1950s Stanford University researcher Carl Dejerassi invented a reliable birth control pill for women. Few could have predicted the eventual consequences of this pharmaceutical invention. These included a sustained decline in the birthrate, vastly expanded workforce participation by women, increased demand for out-of-home child care services, reliance upon television and digital electronics for entertaining children and youth, possible erosion of the two-parent nuclear family (or at least a diminution in the proportion of such units in the overall population), widespread relaxation of prior prohibitions regarding premarital and extramarital sex, and a reduction of the ratio of children to adults in society.

These changes, in turn, contributed to policy demands for publicly subsidized child care, extension of public school kindergarten and preschool programs, new child welfare laws, and added reliance on equal protection laws to ensure women equitable treatment in the workplace. Other effects included proposals for increased regulation of television offerings, a greater openness to overseas immigration to enable the workforce to keep pace with the job demands of a vastly expanding economy, and the increased consumption capacity associated with two income earners in many households.

Examples abound of the reciprocal effects of technological innovation and economic change, and of the combined effect of these two forces upon policy demands. The acceleration of a global economy in the 1990s resulted from the convergence of a number of electronic innovations leading to formation of the Internet and the World Wide Web. These new communication and information transfer technologies liberated capital, creative ideas, and talent from national boundaries. Companies and other organizations, including even international terrorist groups, could now draw upon ideas and recruits from overseas in an unceasing quest for means to lower manufacturing and production costs. The outsourcing of jobs itself reverberated through the United States policy system as electoral candidates and government officials debated and sought means to regulate or limit the flow of offshore jobs.

Ecological conditions can have a profound impact on policy, including education policy. While the intense ecological attention that is presently paid to global warming and carbon emissions has elevated ecological issues into the realm of high politics, there are on-the-ground examples of ecological consequences for schooling. For example, proponents of charter schools saw the policy confusion that followed the 2006 Gulf Coast disasters triggered by Hurricane Katrina as a political opportunity to advocate their cause. The result has been major restructuring of the city's public school system, with regular public schools vastly diminished in authority and enrollment, and the ascendance of charter schools (Levin, Daschbach, & Perry, 2010).

**Preconditions for Macro Policy Change**

Scholars continually probe the primordial ooze from which political actions stem in search of precursors that facilitate or portend significant policy change. Two engaging theories have emerged, both centered on a notion of disequilibrium. When the political system is somehow out of balance, significant policy change is more likely.
Kingdon (1995) stresses the alignment of problems, proposals, and politics, three conditions that conventionally exist independent of one another. A problem occurs when a sufficient proportion of the body politic becomes concerned about an unmet preference or a political demand. In education, there are many proposals for market incentives and arrangements for generating greater competition among schools. Such proposals are often formulated in the absence of a particular problem, proposals, some say, that are in search of a problem to which their proponents can attach them. Politics is the set of circumstances and processes by which, on occasion, problems and proposals find one another.

Kingdon contends that change occurs when political opportunity, the presence of a policy champion, and a set of predetermined solutions that match the problem align. The overwhelming 1964 electoral victory of Democrat Lyndon Baines Johnson over Republican opponent Barry Goldwater for president of the United States supplied Johnson with a huge congressional majority, one so large as to ensure passage of his legislative agenda with hardly a debate. This window of political opportunity facilitated Johnson’s longstanding predisposition to take leadership on issues of poverty and education. He had previously established a blue ribbon task force that created a slate of education ideas. The overwhelming Democratic Party majority rapidly approved these ideas, the principal feature of which was the 1965 ESEA. Johnson had a set of proposals. Having a window of political opportunity enabled him to attach his proposals to a newly defined problem, the eradication or diminution of poverty.

Baumgartner and Jones (1993, 2002) offer a different set of hypotheses regarding dramatic policy redirection. Their concern is with the fact that most policymaking is incremental, controlled by circumstances and by actors connected with the previously mentioned iron triangle of interest groups, bureaucrats, and decision makers. However, from time to time dramatic and deep changes in policy do occur. Baumgartner and Jones label this condition *punctuated equilibrium*—long periods of policy stability interrupted by major alterations to the system. Stability is maintained through the creation of interest group coalitions that can dominate an issue or an already implemented solution to an issue. These dominating positions rely heavily for their sustenance upon the manipulation of policy images, which are generally connected to core political values—equality, liberty, or efficiency. The manipulation of images combines empirical information and popular appeals. These images can be easily communicated and are transmitted directly and simply through campaign commercials and electoral rhetoric.

McDonnell (2007, p. 4) provides the following explanation of the change dynamic espoused by Baumgartner and Jones:

A primary reason changes occur is that those opposed to the policy monopoly or excluded from it constitute slack resources (opposition) policy entrepreneurs can mobilize. They do so through a redefinition of the dominant policy image using ideas that challenge it and capture the imagination of the media, policymakers, and the public. They provide new understandings of policy problems and new ways of conceptualizing solutions. These ideas can fuel powerful changes as they are communicated through a variety of rhetorical mechanisms, including stories about decline or negative consequences resulting from the current policy monopoly and stories of hope about what can be accomplished with a new framing of the policy problem and solution.

**Political Cultures**

The political system has five identifiable cultural components, or self-adopted lenses through which political actors often view reality and attempt to shape it: legislative, regulative, legal, professional, and markets. These political cultures influence the manner in which advocates for a particular change will define a problem or seek predetermined solutions. The unique components of each culture also shape the instruments and processes by which participants attempt to implement and oversee policy.

**Legislative**

This is the culture conventionally associated with policymaking. Deliberative bodies, such as legislatures, city councils, and school boards, confer about issues and render recorded decisions. These politically made rules may be the product of intense controversy characterized by complicated and protracted lobbying by various parties and interests. The activity may have been partisan and linked to identifiable political parties. The newly
enacted policy may alter some existing circumstance, add new service, elevate taxes, or acknowledge some important symbolic issue. Whatever the policy, its construction is characterized by the principal components of the political process: coalition building and bargaining.

**Regulative**

Regulation, too, is a widely perceived component of politics. Many individuals believe that regulations derive singularly from enactments by deliberative bodies. In short, they think of regulations as detailed rules authorized by, and flowing from, relatively abstract policy enactments, something of a bureaucratic follow-on to conventional politics. There is some accuracy to this perception, but by itself it is insufficient.

Executive branch operating agencies also have a political momentum, and sometimes an interest or predisposition of their own. Some of the rules they develop and promulgate stem solely from the momentum of bureaucratic practices and only remotely from explicit political enactments. An organizational imperative seems to exist that specifies that if there is to be a hierarchy of authority, with some individuals enjoying greater decision-making power than others, then there must be rules to enforce the division of labor, specialization of tasks, and an expectation that all similarly situated clients will be treated similarly.

The distinguishing characteristics of the regulatory political culture are codification, rationality, rigidity, specialization, hierarchy, standardization, and efforts to appear objective and independent. Anyone with military experience has encountered a regulatory political culture. Military procedures may eventually be linked to statutory authority, but often one must follow the policy chain of command to very high levels to identify the overarching authority. In fact, the military is quite capable of generating rules all by itself, as are educational institutions. Public universities and school districts have many sets of rules that, however well intentioned and effective, are a product of bureaucratic, and thus not necessarily openly political, processes.

**Legal**

Courts and legal procedures comprise a third major cultural component of the political system. Their identifying features are an ability to frame an issue in keeping with longstanding, judicially sanctioned doctrines, a resort to adversarial techniques, adherence to precedent, appeals to higher authority, and proscribed sets of procedural activities known only to a restricted cadre of certified technicians: lawyers.

**Professional**

The professional culture is the political culture that may be least visible to the general public. From this culture stem procedures and decisions derived from self-reinforcing sets of beliefs of professional participants, not necessarily from court decisions, bureaucratic rules, or political enactments. Distinguishing characteristics are actions taken to restrict entry into the field, protect clients, enhance standards, advance knowledge in a field, and insulate the profession from overtly partisan or selfish interests. Examples include peer evaluation procedures for assessing performance of university professors, peer review among research organizations to determine a manuscript’s publishability or viability of a proposed project, or grading policies for student performance in an institution of higher education or a secondary school.

**Markets**

This political culture is distinguished by a fundamental assumption that clients are sufficiently informed and motivated to operate in their own self-interests and that, in the process of doing so, will, as a collective, promote the public’s long-term interests as well. Market-oriented political cultures certainly have room for rules generated in other spheres. For example, few who believe in deregulating the airline industry would also completely eliminate Federal Aviation Administration rules regarding pilot training or aircraft maintenance. However, the weight of a restrictive argument, from a marketer’s point of view, must be heavily in favor of
some kind of regulation. Further, open choice among services or products, open competition for consumers’ resources, access to capital, and a free flow of information are seen as the sine qua non that should shape policy.

Government Structures

The U.S. Constitution provides the nation with its fundamental governing charter, the collectively affirmed set of rules and structures within which the political process operates. The framers of the Constitution did not envision a national system of education. Indeed, Article I, Section 8 of the U.S. Constitution does not list education among the “enumerated powers” that Congress enjoys, and the 10th Amendment grants states autonomy in virtually all domestic affairs, including education. Moreover, the sovereignty ceded to the states was not dependent on the federal government but instead stems from a state’s citizenry. Consistent with this view, James Madison wrote in The Federalist Papers, first published during 1787 and 1788 (Hamilton, Madison, & Jay, 1961), “The federal and state governments are in fact but different agents and trustees for the people, constituted with different powers, and designed for different purposes” (No. 46, p. 296).

Federal constitutional silence regarding education, coupled with the provisions of the 10th Amendment, means that plenary authority for the formation and operation of the education system rests in the hands of states. At the time states were establishing elementary schools, primitive transportation and communication arrangements rendered centralized governance difficult. Hence, a great deal of authority for schools was initially ceded to local governments. This tradition of local control of schools persists today, though in vastly attenuated form. In fact, whatever the actual degree of local control, the ideology remains powerful symbolically and rhetorically in the conduct of education politics. Thus, even though states are most intensely empowered on matters of education governance and operation, the federal government and local authorities are also involved. This layering renders American education far more complicated structurally and politically than education in virtually any other nation in the world. The overwhelming proportion of other nations, democracies included, relies upon national systems of education.

Vertically Distributed Power in the U.S. Education System

Political scientist Morton Grodzins characterized the governance of U.S. education as “layer cake” federalism, referring to the three governmental strata—local, state, and federal—as the cake layers. One practical expression of the layering is the varying amount of resources contributed by each level of government. The variation is wide across states, but over the last decade state governments paid, on average, between 45 and 50 percent of the expense of public education, local governments paid 40 to 45 percent, and the federal government, as the junior partner, contributed around 10 percent.

Historically, the federal government has taken a permissive role in education. Starting in the mid-20th century, however, a number of events converged to reorient the federal role to a more activist posture. The Servicemen’s Readjustment Act of 1944 (otherwise known as the G.I. Bill) assisted returning veterans in financing their higher education. Also, in 1958, a year after the Soviet Union launched its Sputnik satellite into orbit, Congress enacted the National Defense Education Act, which authorized federal funds for math, science, and language instruction at the precollege level. The 1954 landmark Supreme Court ruling in Brown v. Board of Education and the congressional enactment of the 1964 Civil Rights Act sharpened the federal attention to the needs of disadvantaged students. In addition, the federal government adopted a major antipoverty education program in 1965, Title I of the ESEA.

The expanding federal role notwithstanding, the 50 states continue to assume constitutional and policy authority over much of the domain of public education through state-level education boards. Though far from uniform, selection of state board members is largely by appointment. Governors in 32 states appoint these boards, while elsewhere the process also involves the legislature. Board members in 20 percent of states are elected on either a partisan or nonpartisan basis. Most chief state school officers are appointed, either by their state boards of education (23 states) or their governors (15 states). The rest are chosen in statewide elections.
Despite this variation in state governance structure and political culture (Elazar, 1972; Wirt, 1977), local districts are agencies of their state educational system and exercise only those powers granted to them by the state. States enjoy substantial control over compulsory attendance, accreditation, curriculum, graduation standards, and housekeeping matters such as the school calendar, records, and accounting procedures.

In practice, local governments exercise control over critical resources. Localities can select their own political representatives, decide on fiscal policies, and define the scope of their services. Today, there are over 13,000 local school districts governing nearly 100,000 schools across the nation. Although an overwhelming majority of school boards are popularly elected, those in Baltimore, Boston, Chicago, Cleveland, New York, Philadelphia, and several other cities are appointed either by the mayor or jointly by the mayor and the state’s governor (Wong & Shen, 2013). There are also county systems, principally in the South, that have appointed school boards. Local districts generally maintain control over district organization, instructional personnel, pupil-teacher ratios, staff recruitment, and extracurricular activities. Nonetheless, on school funding and accountability issues, the balance of power has shifted toward greater state control.

**Horizontally Distributed Power in the U.S. Education System**

State level legislative politics play an important role in shaping educational policy because states are central to taxation, expenditure, and constituent concerns, and school legislation often constitutes part of a larger revenue package. Over 80 percent of state education monies are distributed through formulas written by lawmakers; arriving at these formulas involves considerable political give and take and the building of political coalitions.

Coalitions are particularly difficult to create in some circumstances, especially when legislative authority in the education sector is fragmented. In their comparative state study, Rosenthal and Fuhrman (1981) found that education committee chairs quite often do not overlap with the membership on the ways and means and appropriations committees. In each budgetary cycle, education has to compete with other major policy domains, such as welfare assistance, criminal justice, and transportation. Both K–12 and postsecondary education are usually in competition for limited government resources. Given this fragmentation, newly adopted education programs may not be fully funded. Likewise, partisan disagreements between the legislature and the governor’s office, as well as between the two houses within a legislature, can present political obstacles when enacting spending and taxing bills. State boards of education seldom play an important role in this part of the political process.

Conflict is often embedded in urban-suburban-rural divisions within legislatures. An example is Illinois, where the long-standing legislative feud between Chicago and the remainder of the state has been a crucial factor in shaping state revenue and spending policy that affects Chicago schools. Suburban and rural legislators have rarely voted for tax increases to provide additional aid to the Chicago schools. Racial factors may further contribute to these cleavages.

**Influence of Nongovernmental Actors**

Policymakers at the state and district level must also deal with competing demands from multiple organized groups with an interest in educational policy, including, most notably, the two major national teachers unions (McDonnell & Pascal, 1979; Moe, 2002; Loveless, 2001). The American Federation of Teachers (AFT), which has its roots in the labor movement and belongs to the AFL-CIO, is active primarily in urban districts in the northeast and the Midwest. The much larger National Education Association (NEA), whose roots as a professional education association date back to the 19th century, organizes teachers in most other districts.

In his study of the Chicago Teachers Union, an AFT affiliate, Grimshaw (1979) suggested that the AFT typically goes through two phases in its relationship with a local school administration. During the initial organizing phase, when formal bargaining recognition is the objective, the local union largely cooperates with the district and state administration in return for a legitimate role in the policymaking process. Cole (1969) found, for example, that the desire for formal union recognition, where such had never before existed, was a key objective in the 1960 teachers’ strike in New York City. In the second phase, which Grimshaw (1979)
characterizes as “union rule,” the union typically looks to the national union leadership for technical support and political guidance. This support makes it possible for the union to become independent of local and state political leadership and to engage in tough bargaining with the school management over better compensation and working conditions. Consequently, Grimshaw (1979) argued that policymakers “no longer are able to set policy unless the policy is consistent with the union’s objectives” (p. 150).

Policymakers must also contend with an increasingly well-organized taxpaying public, a substantial portion of which no longer has children in the public schools (Kirst & Garms, 1980). The aging population has exacerbated the competition for funds between public education and other sectors such as transportation, public safety, community development, and health care. Discontent with property taxes became widespread during the time of the much-publicized 1978 campaign for Proposition 13 in California. Thirty-nine of the 67 tax or spending limitation measures on state ballots across the nation between 1978 and 1983 were approved (Citrin, 1984). By 2012, 39 states had tax and expenditure limitations in place (Downes & Figlio, this volume). In communities with a substantial representation of retirees, such as those in Florida, referenda for school taxes have met with greater resistance than in settings with larger proportions of households with children.

Since the late 1980s, business-organized lobbying groups have been successful in pushing goals such as higher academic standards, strong accountability measures, and, in some cases, alternative ways of delivering schooling services, including privatization or charter schools. Interesting alliances have emerged surrounding elevated standards and intensified accountability. The Gates Foundation, generally sympathetic with corporate viewpoints regarding education, often finds itself aligned with liberal interests such as the Center for American Progress in an effort to have a more rigorous curriculum and measurement of school and teacher performance.

No Child Left Behind: An Example of Political Dynamics

The NCLB is a massively influential piece of legislation. Its various provisions insinuated the federal government into the warp and woof of American education—all the way into the classroom—and influenced interactions between teachers and students. In addition to its massive practical impact, however, NCLB also illustrates the manner in which abstractions, such as systems theory, iron triangles, policy spheres, policy cultures, policy champions, high politics, and punctuated equilibrium, come together in the reality of education finance and governance.

Consistent with the previously cited writings of Baumgartner and Jones and Kingdon, President George W. Bush took advantage of alterations in global economic and technological conditions to shape a new popular image of American education. He succeeded in using this new image to diminish public sympathy toward the prevailing Clinton Administration interpretation of equality and to promote, in its place, a new concept of equality that furthered the value of efficiency. Bush and his advisers took advantage of a new policy system equilibrium, repurposed efficiency ideas they had long held, and insinuated their enactment into federal policy through NCLB.

Beginnings

NCLB was one of the most prominent proposals of President George W. Bush’s first administration and was submitted to Congress shortly after his initial inauguration in 2001. The act’s principal ideas had been previously honed during Bush’s two terms as governor of Texas, an eight-year administration that had mounted a sustained effort to reform education and had developed a slate of accountability ideas as their major change instrument.

The Bush team trumpeted the need for higher academic standards and more accountability in the education system. In his presidential campaign and early in his administration, Bush discussed the outsourcing of jobs, the trade deficit, and the alarming economic gains of nations such as China. The same kinds of threats had previously been discussed in the 1983 report A Nation at Risk. That report, drafted by the National Commission on Excellence in Education and released by the U.S. Department of Education during the Reagan Administration, marked the beginning of a nationwide focus on academic performance.
The Bush team did not openly disparage equality as a value or policy objective. Rather, it diminished or diluted the previously dominating rhetoric regarding the need for expensive additional services for disadvantaged students, a legacy of the preceding Clinton Administration, and replaced it with the notion that inequality of opportunity was caused not by a lack of educational services, but by the soft prejudice of low expectations. By returning to themes of higher standards and accountability previously proposed in *A Nation at Risk*, Bush and colleagues altered the prevailing policy paradigm. They employed new rhetoric in a manner that was persuasive to the body politic and constructed a new policy attractive to a majority of Congress. In so doing, the Bush Administration punctured the previously prevailing policy paradigm—equality—and replaced it with their efficiency agenda.

**Evolving Politics of Performance-Based Accountability**

The 2001 enactment and subsequent implementation of NCLB represent one of the great political ironies in American history. NCLB is a direct policy descendent of *A Nation at Risk*, which had been promoted by President Ronald Reagan as part of his attempt to reduce the federal government’s role in education. Now, the legislation is unquestionably the most powerful federal influence on American education ever imagined. The twists and turns of political fate are not easily predicted.

At the time *A Nation at Risk* was released, President Reagan was striving to reduce the status and authority of the U.S. Department of Education and thereby shrink the federal role in education. Reagan’s education agenda included enacting tuition tax credits for parents of children attending parochial schools and passing a constitutional amendment that would permit prayer in public schools. Reagan’s Secretary of Education, Terence Bell, did little in practice to push this agenda; however, by 1987, the Reagan Administration had begun to abandon its six-year effort to reduce federal spending in education—a shift that coincided with the appointment of a pragmatic former Tennessee Republican, Senator Howard Baker, as the White House Chief of Staff (Fiske, 1987, p. A14).

Reagan’s drive toward government-wide efficiency and budgetary restraint posed a political challenge to Lyndon Baines Johnson’s Great Society legacy and the powerful policy paradigm that had focused mainly on equity and access. In mediating the tension between efficiency and equity, *A Nation at Risk* proposed an alternative paradigm that focused on quality and standards. The report, however, made no mention of any of the major educational initiatives proposed by the Reagan Administration, nor was it sympathetic with the congressional recommendation for an increase in educational spending. Instead, *A Nation at Risk* left unsettled the question of the amount and types of resources the nation might need in order to reverse the declining trend of its public school system’s educational performance. Rather than terminating the federal role in education, the report called upon the federal government to lead the movement toward “creating a learning society.” *A Nation at Risk*’s recommendations on academic expectations, curriculum, use of time, and teacher quality were featured in all the subsequent legislative changes to the ESEA, including the 1994 Improving America’s Schools Act and the 2001 NCLB (Wong & Nicotera, 2004).

**Post-NCLB Enactment Political Dynamics**

Passage of NCLB in 2001 changed the terms of political engagement among key institutional actors involved in education. The law’s enactment reflected growing public support for holding schools and districts accountable for academic performance, and its accountability-oriented, centrist approach to education drew bipartisan support. Clearly, Republican lawmakers were ready to abandon the 1994 Newt Gingrich-inspired policy platform of dismantling federal involvement. Instead, with NCLB, they chose to substantially broaden federal expectations on outcome-based accountability in public education. In supporting the NCLB, congressional Republicans gave their strong endorsement to a core concern of the Bush Administration, namely raising student achievement. To be sure, Bush was able to bank on Republican control of the presidency and both branches of Congress. He also benefited from the political rapport associated with the first months of a new administration and the widespread public confidence with the federal government in the context of the tragic events of September 11, 2001.

NCLB formally identifies school choice as an option for school restructuring, an indication of relying on the
demand side to drive school improvement. Consistent with the Republican Party’s platform, Bush advocated for greater “consumer” choice during his 2000 presidential campaign, which included school vouchers for private schools. His administration was willing to compromise, however, to ensure annual testing was adopted. As a result, a more limited set of choice arrangements were enacted as part of the provisions for corrective action. When schools fail in consecutive years to meet the Adequate Yearly Progress requirement, students in those schools are granted access to supplementary tutorial service and charter schools. The Obama Administration has continued to support the creation of charter schools as one option for restructuring failing schools.

Challenges to NCLB

Politics has elements in common with physics. For every political action there is something of a reaction, though it may not be equal or precisely opposite. NCLB illustrates this principle. From the moment NCLB was submitted as a bill in Congress, powerful opponents have attempted to prevent its enactment, eviscerate its implementation, and prevent its reauthorization. The opponents, principally, are composed of professional education groups for which the concepts of accountability are threatening. Secondarily, opponents are school administrators who find implementing the massive Act daunting and seek either to mitigate testing and accountability provisions or to delay implementation of the adequate yearly progress schedule. Other criticisms involve the allegation that NCLB, with its insistence upon elevating reading and mathematics test scores, has unduly restricted the school curriculum, and specifically has driven out creative subjects such as the arts and music. In addition, critics contend that the Act has such a large number of underdefined components (e.g., that every child deserves a highly qualified teacher) that federal regulations outstrip practical or research knowledge. It is not known, for example, what in reality comprises a highly qualified teacher.

In the face of such local and state resistance, the U.S. Department of Education initially relaxed some requirements on a case-by-case basis. For example, Chicago succeeded in gaining federal approval to provide its own after-school tutoring programs for students in schools that failed the adequate yearly progress standards rather than rely on outside vendors as prescribed by the legislation. In return for permission to continue its supplemental services, the city agreed to reduce barriers for private vendors to provide tutorial services. Similar waivers were subsequently granted in New York City, Boston, Memphis, and elsewhere.

Barack Obama’s 2008 presidential election enabled his appointees to further influence federal education policy. By 2009, congressional consensus over education policy, prominent in the prior Bush Administration’s enactment and implementation efforts, had begun to dissolve. In the absence of political agreement and the failure to reauthorize the ESEA, the Obama Administration creatively resorted to heavy reliance upon executive branch authority—an example of a regulative political culture—and induced states to intensify and expand accountability procedures through the offering of waivers exempting states from several outmoded NCLB provisions. In addition, the Obama Administration used legislation designed to address the effects of the 2008 Great Recession to expand federal education funding and to provide the education secretary with $5 billion in state incentive funding directed at higher academic standards, wider accountability, and expansion of performance measurement, a remarkable addition to the executive branch arsenal for inducing state change.

Conclusion

By using the NCLB as an illustration, this chapter explains the multilayered governmental structure fostered by the U.S. Constitution and exemplifies the interaction of politics and government. Deeper understanding of these various interactions may enable policymakers, eventually, to tailor policies more to public preferences and, possibly, more forcefully contribute to the public’s wellbeing.

Notes
. Only health care provision and government debt annually involve larger government appropriations.

. The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people.

. The Bush campaign’s attention to education came at a time when public confidence in schools had been declining for more than a 30-year period. In 1977, for example, 53 percent of the public revealed a high level of confidence with public education. By May 2005, the confidence level fell to only 37 percent (Public Agenda, 2006). However, the public seemed not ready to give up reforming the existing system of public schools. According to the Gallup Poll conducted in the summer of 2003, 73 percent of survey respondents wanted to reform the existing public school system, while only 25 percent preferred alternative approaches such as school vouchers.

. A reduced federal government role in education reappeared as a policy objective in 1994 under the leadership of Republican House Speaker, Newt Gingrich, of Georgia. However, Gingrich’s agenda was never fulfilled prior to his ouster as Speaker.

References


Chicago: University of Chicago Press.
Contemporary education policy stresses accountability but defines the goals of public education narrowly, usually holding schools and teachers accountable primarily for adequate standardized test scores in mathematics and reading. Yet because public schools provide collective benefits, such as a common set of democratic values, Americans have historically defined schools’ goals more expansively.

As this chapter shows, a careful review of the history of American educational thought demonstrates that policymakers have always believed that schools should be held accountable for students having sufficient background in history and in world and civic affairs to exercise the rights and responsibilities of citizenship, and for students having the organizational skills and proficiency in decision making and cooperative behavior that will allow them to participate thoughtfully in the democratic processes of their communities and nation. Americans want students to have sufficient appreciation of literature and of the visual, musical, and performing arts to be able to engage in fulfilling adult leisure activities. Policymakers have regularly held that schools should also graduate students with habits of intellectual curiosity, creative imagination, and personal discipline—patience, persistence, and self-confidence—which enhance both workplace productivity and personal fulfillment. Graduates should know how to resolve personal and social conflicts peaceably and in a manner that reconciles competing interests with justice and fairness. Youths should graduate with a broad knowledge of science—not only to choose technical careers if they wish, but also because public problems on which citizens in a democracy must deliberate require scientific insight for intelligent resolution. Further, schools are expected to inspire young people to develop habits of exercise, hygiene, and nutrition and to lead healthy and responsible teen and adult lives.

The contemporary near-exclusive emphasis on basic academic skills is an historic aberration. Widely publicized contemporary surveys show that Americans do expect students to graduate with the literary and mathematical skills that enable them to enroll in college or technical school or to succeed in vocations of their choice. Yet these surveys also find support for the idea that assessment and accountability systems should include more than these skills alone (e.g., Rose & Gallup, 2006).

Pre-Twentieth Century Goals of Education

When the Founding Fathers called for public education, their motives were mostly political. Instruction in reading was the way to teach citizens to make wise political decisions. History instruction was thought to teach students good judgment, to help them learn from prior generations’ mistakes and successes, and to inspire character traits such as honesty, integrity, and compassion. The Founders had no doubt that education could produce good character, and it would never have occurred to them that reading and arithmetic alone would guarantee good citizenship.

In 1749, Benjamin Franklin proposed that Pennsylvania establish a public school that should, he said, “place as much emphasis on physical as on intellectual fitness because ‘exercise invigorates the soul as well as the body.’” As for academics, Franklin thought history particularly important, because “questions of right and wrong, justice and injustice, will naturally arise” as students debate historical issues “in conversation and in writing.” They should also discuss current controversies in order to develop logic and reasoning (Franklin,
George Washington’s goals for public schools were also political and moral. In his first message to Congress, he advocated public schools that would teach students “to value their own rights” and “to distinguish between oppression and the necessary exercise of lawful authority” (Washington, 1790, para. 12). His farewell address warned that because public opinion influences policy in a democracy, “it is essential that public opinion should be enlightened” by schools teaching virtue and morality. He wanted to go further, but his speech-writer (Alexander Hamilton) cut from the address a proposal for a national public university that would encourage tolerance of diversity: integrating students of different backgrounds to show them there is no basis for “jealously and prejudices” (Washington, 1796, p. 154).

Thomas Jefferson thought universal public education was needed primarily to prepare voters to exercise wise judgment. He wanted not what we now call civics education—learning how bills are passed, how long a president’s term is—rather, he thought schools could prepare voters to think critically about candidates and choose wisely (Jefferson, 1818; see also Wagoner, 2004).

In 1787, the Northwest Ordinance provided funds for new states to establish public schools. To enter the union, states were required to adopt a commitment to public education. Indiana’s 1816 constitution, for example, asserted that education is “essential to the preservation of a free government,” and instructed the legislature to establish schools and to enact other policies to promote “arts, sciences, commerce, manufacture and natural history; and ... countenance and encourage the principles of humanity, honesty, industry and morality” (Thorpe, 1909, p. 1069).

Early 19th century labor unions insisted that public schools for the poor should include not only basic reading and arithmetic, but the more important intellectual development found routinely in schools for the wealthy. An 1830 workingmen’s committee, for example, examined Pennsylvania’s urban public schools. The committee denounced urban schools for instruction that “extends [no] further than a tolerable proficiency in reading, writing, and arithmetic ...” Equality, the committee concluded, is but “an empty shadow” if poor children do not get an “equal education .... in the habits, in the manners, and in the feelings of the community” (as cited in Cremin, 1951, p. 51).

In 1837, Horace Mann was elected secretary of the newly created Massachusetts Board of Education and thereafter wrote 12 annual reports to encourage support for public schools. One report stressed the importance of vocal music, not only because musical notes have mathematical relationships, but because it was a social equalizer and encouraged pacific and unifying sentiments—patriotism, for example (Mann, 1845). Another of Mann’s reports, following a visit to Europe, concluded that education in reading and arithmetic did not alone ensure democratic values. Prussian students were literate, after all, but supported autocracy. Mann concluded that Massachusetts schools should not be held accountable for academics alone but must inculcate democratic moral and political values so that literacy will not be misused (Mann, 1844).

**Twentieth Century Views**

As schooling expanded in the 20th century, the federal Bureau of Education commissioned a 1918 report, *The Cardinal Principles of Secondary Education*. Although some contemporary historians have promoted the notion that the *Cardinal Principals* turned American education away from academic skills (e.g., Ravitch, 2000), this is an exaggeration. In fact, the document asserted that “much of the energy of the elementary school is properly devoted to teaching certain fundamental processes, such as reading, writing, arithmetical computations, and the elements of oral and written expression” and that secondary schools should be devoted to applying these skills. But the document argued that academics were not enough; continuing in the tradition of the Founding Fathers and Horace Mann, it urged a balanced approach to education goals.

As its first goal, the committee listed physical activity and instruction in personal hygiene and in public health. Its second goal was academic skills. Third was preparation for traditional household roles of husbands and wives. Fourth was vocational education, including selection of jobs appropriate to each student’s abilities and interests, and maintenance of good relationships with fellow workers.

Like the Founders, the committee emphasized in its fifth goal the need for civic education: preparation for participation in neighborhoods, towns or cities, states, and nation. *The Cardinal Principals* devoted more space
to civic education than to any other goal, stressing that schools should teach “good judgment” in political matters and that students can learn democratic habits only if classrooms and schools are run with democratic methods. Even the study of literature should "kindle social ideals."

The sixth goal was "worthy use of leisure," or student appreciation of literature, art, and music. And lastly, the seventh goal, ethical character, was described as paramount in a democratic society. It included developing a sense of personal responsibility, initiative, and the "spirit of service" (Commission on the Reorganization of Secondary Education, 1918. p. 15).

Two decades later, the National Education Association (NEA), then a quasi-governmental group that included not only teachers but all professionals and policymakers in education, considered how schools should respond to the Great Depression. The NEA warned in 1938: "[T]he safety of democracy will not be assured merely by making education universal" (in other words, simply by making all Americans literate). "The task is not so easy as that. The dictatorships [Germany, Italy, Japan, and the Soviet Union] have universal schooling and use this very means to prevent the spread of democratic doctrines and institutions" (Educational Policies Commission, 1938). Schools, it went on, should also teach morality: justice and fair dealing, honesty, truthfulness, maintenance of group understandings, proper respect for authority, tolerance and respect for others, habits of cooperation, and work habits such as industry and self-control, along with endurance and physical strength.

Prefiguring our contemporary dilemmas, the 1938 report went on to warn that tests measuring only "basic skills may recognize objectives of education which are relatively unimportant." Rather, we should evaluate schools by measures such as whether graduates are "sympathetic in the presence of suffering and indignant in the presence of injustice," show "greater concern about questions of civic, social, and economic importance," are "living in accordance with the rules of health," and appreciate "their rich inheritance in art, literature, and music" (pp. 15–16).

Another two decades had passed when the Rockefeller Brothers Fund convened leaders to make public policy recommendations. Nelson Rockefeller (subsequently New York’s governor and Gerald Ford’s vice president) chaired the project; Henry Kissinger (later secretary of state) was staff director. The 1958 Rockefeller Report asserted: "Our conception of excellence must embrace many kinds of achievement.... There is excellence in abstract intellectual activity, in art, in music, in managerial activities, in crafts-manship, in human relations, in technical work." Testing was important for sorting future scientists and leaders. But, the panel warned that "[d]ecisions based on test scores must be made with the awareness of the .... qualities of character that are a necessary ingredient of great performance [.... aspiration or purpose.... courage, vitality or determination] (Rockefeller Brothers Fund, 1958, p. 29).

For the last 40 years, litigants have argued that states are obliged to finance an "adequate" education and state courts have had to define what this means. True to American traditions, courts have proposed definitions that went beyond math and reading test scores alone.

The earliest such decision was issued by the New Jersey Supreme Court, finding, in 1973, a constitutional requirement "to equip a child for his role as a citizen and as a competitor in the labor market" (Robinson v. Cahill, 1973, p. 515). The Court later elaborated, stating,

Thorough and efficient means more than teaching.... skills ... It means being able to fulfill one’s role as a citizen, a role that encompasses far more than merely registering to vote. It means the ability to participate fully in society, in the life of one’s community, the ability to appreciate music, art, and literature, and the ability to share all of that with friends....


These are goals sought by wealthy districts, the court said, and must be pursued in low-income areas as well (Abbott v. Burke II, 1990).

Efforts to Synthesize Goals: 1950s to 1990s

This chapter is not the first to review goals of American education and conclude that an excessive stress on basic academic skills violates a national consensus. In the late 1950s, a University of Chicago team synthesized all education goals it could identify that had previously been embraced by Americans. Led by Lawrence...
Downey, the researchers polled 15 representative communities in the nation and in Canada, surveying educators and members of service, social, labor, management, church, ethnic, and racial-justice organizations. Nearly 1,300 educators and 2,500 noneducators were asked to sort cards on which 16 alternative education goals were described, and place these cards in order of relative importance.

For elementary schools, educators ranked “intellectual skills” (defined as the skill to acquire and transmit information, as opposed to the possession of knowledge) as the highest priority. Next was development of a desire for knowledge, or “love of learning.” Third was students’ ability to cooperate with others in day-to-day life. Fourth was creativity, fifth was moral integrity, sixth was good citizenship, and seventh was emotional health and stability. Nine other goals followed, with lesser importance. Top priorities were similar for high schools. The noneducator community leaders established similar, though not identical priorities (Downey, 1960).

In 1975, John Goodlad commenced a review of American public education that culminated in his 1984 book, A Place Called School. Analyzing goals that had been embraced by state and local boards of education and commissions that had studied education, Goodlad defined a consensus on the goals of education that included basic academic skills; vocational preparation (including positive attitudes towards work and habits of good workmanship); problem solving skills and intellectual curiosity; interpersonal skills such as respect, trust, and cooperation; citizenship, including a willingness to participate in national and community political life; emotional and physical well-being; and moral and ethical character, including a commitment to “strengthen the moral fabric of society” (Goodlad, 1994, p. 51 and also in Goodlad, 1984).

As demands have recently grown for school accountability based on test scores, some scholars and officials have sounded alarms about the potential for high-stakes basic skill tests to distort curricula. In 1987, a committee of the National Academy of Education issued this warning (p. 64):

At root here is a fundamental dilemma. Those personal qualities that we hold dear—resilience and courage in the face of stress, a sense of craft in our work, a commitment to justice and caring in our social relationships, a dedication to advancing the public good in our communal life—are exceedingly difficult to assess. And so, unfortunately, we are apt to measure what we can, and eventually come to value what is measured over what is left unmeasured. The shift is subtle, and occurs gradually. It first invades our language and then slowly begins to dominate our thinking. It is all around us, and we too are a part of it. In neither academic nor popular discourse about schools does one find nowadays much reference to the important human qualities noted above. The language of academic achievement tests has become the primary rhetoric of schooling.

Citing this warning, the Department of Education’s National Center for Education Statistics (NCES) proposed, in 1991, a national indicator system that reflected this balance of education goals. NCES urged that, in addition to academic competence in core subjects, measured “learner outcomes” should include tolerance, comprehending pluralism, self-direction, responsibility, commitment to craft, and other measures (Morgan, 1991). Subsequently, NCES commissioned the design of a variety of “input, output, outcome, and efficiency” indicators. The American Institutes for Research produced an inventory of such indicators (Peternick et al., 1999), but did not define their relative importance, and NCES never proceeded to develop such a system. As the testing frenzy grew in the 1990s, the intent was forgotten.

Current Consensus on the Goals of Education

This chapter’s authors attempted, in 2005, a new synthesis of education goals that had been established through 250 years of American history. We defined eight broad goal areas that seemed to be prominent in each era, although certainly emphases changed from generation to generation. We then presented these goals to representative samples of American adults, of school board members, state legislators, and school superintendents and asked respondents to assign a relative importance to each of the goal areas. (For the full survey methodology and detailed results, see Rothstein, Jacobsen, & Wilder, 2008.) Respondents recorded their weightings via an online survey. Average responses of all adults, board members, legislators, and superintendents were very similar and are thus combined and presented in Table 5.1. As our results demonstrate, the representative samples we polled stressed a balanced, though not equally weighted set of outcomes.

Table 5.1 The Relative Importance of the Goals of Education

83
Even after two decades of state and federal policies that have disproportionately emphasized accountability for reading and mathematics proficiency, surveys of parents and members of the public continue to find results similar to those we present above. In fact, many parents now express concern about the exclusive focus on reading and math test scores. Public Agenda, a nonpartisan research organization that often examines education, conducted several surveys of parental attitudes towards testing and curricular goals. In its 2013 report, Will It Be on the Test, Public Agenda finds that parents believe there are many factors including, but not exclusively, standardized test scores, that ought to be considered when judging whether schools are effective (Johnson, 2013). While parents expressed that standardized test scores can provide some useful information, many believe too much emphasis on testing distorts student learning. In a separate study that included discussions with secondary school parents, Public Agenda found parents wanted more emphasis on fine arts (40 percent of parents), civics and social studies (45 percent), and teaching students to work well as part of a team (62 percent) (Johnson, Rochkind, & Ott, 2010). Similarly, the Thomas B. Fordham Institute found in its national study of K–12 parents that some variation existed in parental preferences. Parents emphasized a range of goals: 36 percent wanted schools to offer "vocational classes or job-related programs," 24 percent expected schools to emphasize "instruction in citizenship, democracy and leadership," and 22 percent wanted students to learn "how to work with people from diverse backgrounds," while only 23 percent ranked highly the desire to see a school have "high test scores" (Zeehandelaar & Winkler, 2013).

Contemporary Goal Displacement

Today, federal and state accountability policies (like No Child Left Behind and Race to the Top) distort public education’s widely shared goals. By basing sanctions heavily on math and reading test scores, the law creates incentives to narrow education goals by limiting time spent on other important curricular objectives. Research has shown that this reorientation of instruction disproportionately affects low-income and minority children, so achievement gaps may actually widen in domains for which schools are not now held accountable.

The Center on Education Policy has tracked the implementation of NCLB since 2002 and found that from 2001 to 2007, over half (58 percent) of the elementary schools surveyed increased time allocated for English language arts instruction, by an average of 141 minutes per week. Similarly, 45 percent of schools increased

<table>
<thead>
<tr>
<th>Goal area</th>
<th>Relative importance (weights, in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basic Academic Skills in Core Subjects</strong></td>
<td>22</td>
</tr>
<tr>
<td>Reading, writing, math, knowledge of science and history.</td>
<td></td>
</tr>
<tr>
<td><strong>Critical Thinking and Problem Solving</strong></td>
<td>18</td>
</tr>
<tr>
<td>Able to analyze and interpret information, use computers to develop knowledge, apply ideas to new situations.</td>
<td></td>
</tr>
<tr>
<td><strong>Social Skills and Work Ethic</strong></td>
<td>12</td>
</tr>
<tr>
<td>Good communication skills, personal responsibility, the ability to get along well with others, and work with others from different backgrounds.</td>
<td></td>
</tr>
<tr>
<td><strong>Citizenship and Community Responsibility</strong></td>
<td>11</td>
</tr>
<tr>
<td>Know how government works, how to participate in civic activities like voting, volunteering, and becoming active in communities.</td>
<td></td>
</tr>
<tr>
<td><strong>Preparation for Skilled Work</strong></td>
<td>10</td>
</tr>
<tr>
<td>Vocational, career, and technical education that will qualify youth for skilled employment that does not require a college degree.</td>
<td></td>
</tr>
<tr>
<td><strong>Physical Health</strong></td>
<td>9</td>
</tr>
<tr>
<td>A foundation for lifelong physical health, including good habits of exercise and nutrition.</td>
<td></td>
</tr>
<tr>
<td><strong>Emotional Health</strong></td>
<td>9</td>
</tr>
<tr>
<td>Tools to develop self-confidence, respect for others, and the ability to resist peer pressure to engage in irresponsible personal behavior.</td>
<td></td>
</tr>
<tr>
<td><strong>The Arts and Literature</strong></td>
<td>9</td>
</tr>
<tr>
<td>Participate in and appreciate the musical, visual, and performing arts and develop a love of literature.</td>
<td></td>
</tr>
</tbody>
</table>

1 The weights shown are a simple average of the average responses for each of the four surveyed groups.
instructional time for mathematics by an average of 89 minutes per week. Because the school day did not lengthen, these increases cut time from social studies, science, arts/music, recess, and physical education. Each of these areas was reduced, on average, by about one-third (Center on Education Policy, 2008). Moreover, these curricular changes were not uniformly distributed. A 2005 survey by the same organization found that 97 percent of high-poverty districts had new minimum time requirements for reading, while only 55 percent of more affluent districts had them (Center on Education Policy, 2005). Thus, although NCLB aims to narrow achievement gaps in math and reading, its unintended consequence is likely to widen gaps in other curricular areas.

Getting Accountability Back on Track

While education policymakers often claim that accountability will promote improved academic performance, accountability should also improve public oversight of the schools and restore trust in public education by making its accomplishments transparent. Addressing declining public satisfaction with public education is critical as confidence has declined significantly since the 1970s. In 2013, nearly a quarter of the public expressed "very little" confidence in the U.S. education system (Gallup Poll, 2013). But today’s reporting policies fail to cover the wide range of expectations the public holds for its schools, and therefore may have a distorted impact on public satisfaction. Satisfaction results from the interaction of one’s expectations of a good or service and one’s experiences with the actual good or service: expectations influence how one judges experiences (Oliver, 1980). Even if people agree on the absolute performance level of a good or service, they may express differing levels of satisfaction due to different expectations.

Contemporary dissatisfaction with schools probably results, in part, from reports of school performance that attempt to influence expectations as well as actual reported results. This arises because federal, state, and local governments now report math and reading scores by "proficiency" standards. The public cannot compare scale scores to its own expectations, but rather only has information about how scores compare with governmentally established expectations.

However, increased dissatisfaction may also result from the failure of accountability systems to incorporate all of the public’s expectations, not only those for math and reading performance. To improve satisfaction, policymakers and school leaders must demonstrate that schools are meeting other expectations as well. Improved test scores alone will not meet the public’s goals. What would an accountability system look like if it created incentives for schools to pursue a balanced set of goals?

It would certainly include standardized tests of academic skills. Other standardized tests should be added. For example, tests of physical fitness (such as upper body strength) and simple measures of body weight can be standardized to shed light on the efficacy of schools’ physical education programs. Schools that sacrifice essential physical education for excessive math and reading drills would lose incentives for such practices with accountability so structured.

Balanced accountability should also utilize measures that are more difficult to standardize but equally valid. Student writing and analysis of contemporary issues, and student performances in the arts, scientific experimentation, and debates, should also be included. School accountability does not require such assessments of each student every year, but only of a random sample drawn periodically.

Accountability also requires less immediate measures that nonetheless reflect on school adequacy. As most states have adopted unique student identifier systems, it has become possible to link data from surveys of representative groups of young adults to the efficacy of schools they attended. By such means school civics programs can be judged by whether graduates register and vote, participate as community volunteers, or contribute to charity. Survey data can indicate the adequacy of students’ literacy instruction by the standard of whether, as young adults, they read for pleasure. Surveys can also indicate the adequacy of students’ physical education in part by whether, as young adults, they exercise regularly.

Balanced accountability also requires school inspections, differing from today’s accreditation procedures. Inspection teams should include, in addition to professional educators, representatives of elected officials, businesses, labor, and community groups. Inspectors should judge whether students are engaged in group activities likely to develop the teamwork so valued by employers, or whether classroom discussions aim to
development critical thinking that leads to intelligent voting.

We are a long way from developing an accountability system that is true to American traditions and to our contemporary goals for public schools. We could move towards such a system, but test-based accountability is taking us in the opposite direction.

Acknowledgments

Research for this chapter was supported by the Teachers College Campaign for Educational Equity. Views expressed in this chapter are those of the authors alone and do not necessarily represent positions of Teachers College, Columbia University.

References

Mann, H. (1845). Eighth annual report of the Board of Education together with the eighth annual report of the Secretary of the Board. Boston: Dutton and Wentworth.
Academy of Education.


6
Quantitative Research Methods in Education Finance and Policy

PATRICK J. McEWAN

Introduction

Researchers in education finance and policy rely on a myriad of quantitative methods. The most common include regression analysis, a mainstay of social science research, but they also include experimental and quasi-experimental methods (Murnane & Willett, 2011; Shadish, Cook, & Campbell, 2002). These methods are particularly suited to addressing research questions about the causal relationship between a policy or program and education outcomes. Do school finance reforms increase the equity of school expenditures? Does attending a private instead of a public school improve students’ achievement? Does financial aid increase the probability that students attend college? This chapter describes a range of quantitative methods that can be used to address causal questions, placing special emphasis on each method’s rationale, intuition, and pitfalls.

Methods for What?

Defining Research

In 2002, following a request from the U.S. Department of Education, the National Research Council (NRC) defined principles of scientific inquiry in education research (Shavelson & Towne, 2002). To paraphrase the NRC report, education research should: (1) pose important questions; (2) investigate them empirically, with appropriate research methods; (3) provide an explicit answer, after assessing and discarding plausible alternatives; (4) replicate and generalize the answers across contexts; and (5) do so in a transparent and professionally accountable way. The greatest controversy lies in the definition of “important” research questions, which shape views on “appropriate” methods. Shavelson and Towne (2002) define three categories of questions: (1) descriptive (what is happening?), (2) causal (does it work?), and (3) process (how does it work?).

Descriptive research establishes or refutes patterns in the data, inspires theoretical explanations of the observed facts, guides the design of causal research, and provides better context for interpreting and generalizing causal results. Causal research tests for cause-and-effect relationships, rather than mere correlations, between policy interventions and policy outcomes. It helps test, refine, and possibly discard theoretical explanations of empirical regularities in education. Most practically, it helps policymakers assess the relative merits of interventions and allocate resources to better ones, however defined. Process research inquires why a policy intervention does or does not affect policy outcomes. Did a single component of the intervention not work? Was it implemented well or poorly? Did teachers, students, or other participants respond by altering their behavior? To the extent that it unpacks the mechanisms explaining a “black box” causal effect, process research helps establish whether the same result can be replicated elsewhere.

Choosing Research Methods
The choice of a research method hinges on the question posed. Descriptive research employs standard quantitative techniques (Tukey, 1977) to describe the central tendency and dispersion of single variables (e.g., school expenditures or teacher characteristics) and their statistical association with other variables (e.g., student characteristics or test scores). Descriptive research also employs qualitative techniques—including case studies, ethnographies, and pragmatic reportage—to describe classrooms, schools, legislation, and political contexts.

Causal research uses mostly quantitative techniques, especially variants of regression analysis, to isolate the unique contribution of a policy to an education outcome. Regression methods are increasingly combined with, or supplanted by, experimental or quasi-experimental research methods, described below. Causal research employs qualitative methods less commonly, despite the potential applications (King, Keohane, & Verba, 1994).

Finally, process research is conducted with mixed, but frequently qualitative methods. In the best of cases, process research is built into an existing causal research design. For example, researchers have employed qualitative methods, especially teacher interviews and classroom observations, to explain the large or small effects of reduced class sizes (Zahorik et al., 2000), Catholic school attendance (Bryk, Lee, & Holland, 1993), and whole-school reform (Cook et al., 1999; Cook et al., 2000).

**The Importance of Causal Research**

This chapter emphasizes quantitative methods for causal inference. In particular, it discusses regression analysis and related statistical techniques, as well as popular experimental and quasi-experimental research designs for analyzing the causal impact of policy interventions. It does so because these methods are foundational in modern research on education finance and policy, as evidenced by other chapters in the Handbook. Even so, some important methodological innovations have been slow to filter down to the day-today practice of education policy researchers, in part because the methods are confined to specialized journal articles.

At least two factors have spurred the interest in causal research. First, policymakers increasingly demand evidence that education policies or programs have been rigorously shown to cause improvements in student outcomes. This is partly manifested in the increased federal support for high-quality causal research, channeled through the Institute for Education Sciences. Relatedly, the What Works Clearinghouse reviews education research and judges its ability to derive valid causal inferences about the impact of education programs (What Works Clearinghouse, 2013).

Second, education researchers have access to increasingly rich education data that lowers the cost of research and facilitates the application of better methods (Loeb & Strunk, 2003; Perez & Socias, 2010). The data fall into three broad categories: (1) national samples of students; (2) state and local administrative data that include repeated observations on a population of students; and (3) site-specific data gathered by researchers for field experiments. The second and third categories have been a particular catalyst to recent education policy research, because they facilitate better methods described below.

**What This Chapter Does Not Address**

Given the chapter’s limited scope, it does not address three issues. First, the chapter does not discuss the application of quantitative methods to descriptive- or process-oriented research questions. Second, the chapter does not describe qualitative methods, despite their evident or potential merits in addressing descriptive, causal, or process questions. Third, it does not discuss a few quantitative methods that, while used by policy researchers, are the subject of their own specialized literatures. These include cost-benefit and cost-effectiveness analysis (see Belfield, this volume, as well as Levin & McEwan, 2001; McEwan, 2012), meta-analytic techniques for summarizing the results of many studies (Cooper, Hedges, & Valentine, 2009), and methods of measuring and comparing student achievement (Crocker & Algina, 1986; Kolen & Brennan, 2004).
Definitions of Research Terms

The results of cause-testing research, regardless of the method, are commonly judged by two criteria: *internal validity* and *external validity* (Cook & Campbell, 1979; Meyer, 1995; Shadish et al., 2002). A result is internally valid when it identifies a believable causal link between a policy or program and an education outcome. (The policy or program is often generically called a *treatment*.) A causal result is externally valid when it can be generalized to modified versions of the policy treatment, to alternate measures of student outcomes, to diverse populations of students or schools, or to different policy contexts. Much of the following discussion emphasizes the importance of improving internal validity (Barrow & Rouse, 2005). However, an internally valid result may be of little use to policymakers if it cannot be usefully generalized.

To take one example, California’s large-scale and costly class size reduction was inspired by small-scale research in Tennessee that demonstrated positive—and internally valid—effects on student achievement (Mosteller, 1995; Krueger, 1999). Unlike the Tennessee research, California’s large-scale implementation substantially increased demand for teachers. Statewide, deteriorating teacher quality appears to have offset the positive effects of class size reduction (Jepsen & Rivkin, 2009). The California experience highlights the trade-offs that often exist between internal and external validity.

The causal effect of a policy treatment is the difference between students’ outcomes when treated, and the same students’ outcomes when not treated. The latter is simply called the *counterfactual*. Short of procuring a time machine, it cannot be observed because treatments cannot be undone. Instead, research methods are employed to “create reasonable approximations to the physically impossible counterfactual” (Shadish et al., 2002, p. 5).

Researchers typically estimate counterfactual outcomes by identifying a separate group of untreated students, called a *control group* or *comparison group* (the former term is usually reserved for randomized experiments, but is used loosely). Members of the control or comparison group should be similar to their treated counterparts, on average, in every respect but for exposure to the treatment. In practice, the groups are often dissimilar, in ways that affect the outcomes of interest but have nothing to do with the treatment (for instance, students’ families have different incomes, because higher-income families were more or less likely to choose the treatment). If this occurs, then the mean outcomes of comparison students incorrectly estimate the policy counterfactual. Thus, a simple difference in the average outcomes of treatment and comparison students can yield a misleading estimate of the treatment’s causal effect, and is said to suffer from *selection bias*. Selection bias is a pervasive threat to internal validity, and the research methods discussed below are aimed at lessening it (on other threats, see Shadish et al., 2002; Meyer, 1995).

Researchers use two broad methods, sometimes in combination, to ensure that members of treatment and control/comparison groups are similar, on average. First, they make statistical controls for observed differences between students, using regression analysis and/or propensity score methods. Second, they influence how students are assigned to treatment and control groups. Cause-testing research is often lumped into three broad categories: *experimental*, *quasi-experimental*, and *nonexperimental*. The essential difference among categories is the degree of control the researcher exerts over who is assigned to the policy treatment (whether students, teachers, schools, districts, or states), and who is assigned to the control/comparison group.

In experiments—often referred to as randomized, controlled trials (RCTs)—assignment is determined by luck of the draw, as in a researcher’s flip of a coin. In quasi-experiments, the broadest category of research, assignment may contain elements of randomness; deliberate, formula-based assignment by researchers or policymakers; or the individual choices of students, parents, or administrators (called *selection*). In nonexperimental research, the researcher exerts no influence, and assignment is entirely due to selection. When the researcher exerts greater control, then causal results often possess greater internal validity. The following sections develop this point.

Methods for Answering Causal Questions

*Statistical Controls for Observed Variables*
Suppose that researchers collect nonexperimental data from students who attend either private or public schools but were not encouraged or coerced to do so by researchers. The causal question is whether attending private school improves test scores, on average. A naïve researcher would simply estimate the difference between the average test scores of private school students and public school students, and ascribe it to the causal effect of school type. But the difference could be explained by preexisting differences in students that are the result of selection. For one, private students in tuition-paying schools have higher incomes, on average, which might be associated with higher test scores.

**Regression Analysis.** Regression analysis is the first line of defense against this kind of selection bias. A linear regression model can be written as:

\[ A_i = \beta_0 + \beta_1 P_i + \beta_2 X_i + \epsilon_i \]

where

- \( A \) represents the test score of the \( i \)th student in the sample (the subscript \( i \) might range from 1 to 1000)
- \( P \) indicates whether a student attends a private school (\( P = 1 \)) or public school (\( P = 0 \))
- \( X \) indicates the value of a control variable, like family income, that one wishes to hold constant
- \( \epsilon \) is an error term unique to each student (The error term captures the notion that test scores vary, for unobserved reasons, even among students attending the same school type and with the same incomes.)

Using the method of ordinary least squares, researchers obtain estimates of \( \beta_0, \beta_1, \) and \( \beta_2 \). In the absence of controls for \( X \), the estimate of \( \beta_1 \) would be interpreted as the average difference between private and public students’ achievement (the naïve estimate from before). Upon controlling for \( X \), it is the average difference holding constant family income. The immediate question is whether the difference can now be interpreted as the causal effect of private school attendance.

The causal interpretation rests on an assumption of regression analysis: that private school attendance \( (P) \), controlling for \( X \), is uncorrelated with positive or negative shocks in test scores captured in the error term, \( \epsilon \). What could produce such correlations? Suppose the existence of an unmeasured variable, \( M \), that gauges parent motivation. Further suppose that children of motivated parents disproportionately attend private schools \( (M \) and \( P \) are positively correlated, even controlling for \( X \)) and that the children of motivated parents obtain higher test scores \( (M \) and \( A \) are positively correlated, even controlling for \( X \)).

In regressions that do not control for \( M \), the net result is that attending private schools tends to be accompanied by positive shocks in students’ test scores, the (noncausal) influence of greater unobserved motivation among their parents. In this simple example, estimates of the coefficient \( \beta_1 \) would be too large because of selection bias, leading to overly optimistic conclusions about private school effects. Yet, selection bias could also work in the opposite direction, depending on the sign of partial correlations between the omitted variable(s), the dependent variable \( (A) \), and the key treatment variable \( (P) \). Omitting variables such as \( M \) creates no bias in estimates of \( \beta_1 \) if (1) the omitted variables are uncorrelated with \( A \), and/or (2) the omitted variables are uncorrelated with \( P \).

In nonexperimental research settings such as this, researchers have few remaining options. One is to gather additional variables, but that is cold comfort to users of secondary data. Even controlling for hundreds of variables proves unconvincing in many contexts. For example, the archetypal nonexperimental study in education policy regresses student test scores on student, family, and school variables. The third category often includes a key policy variable such as class size. Yet, test scores (say, at the end of sixth grade) are the cumulative product of family and school inputs received by students from birth onwards (Todd & Wolpin, 2003). Data sets, even detailed longitudinal ones, can never fully and accurately measure all inputs. Variables such as class size are probably correlated with unobserved variables that determine achievement, and estimates of their causal effects are unpredictably biased upward or downward. For example, Urquiola (2006) shows that disadvantaged, rural students in Bolivia, by virtue of their location, are more likely to attend smaller classes. Presuming that some features of “disadvantage” are unobserved and lead to lower test scores, regression-based estimates of the causal effect of small-class treatments are likely biased towards finding no effect.

In test score regressions, researchers often resort to controlling for test score measurements taken at earlier moments in students’ careers (say, at the beginning of sixth grade). By controlling for pretests in so-called
value-added regressions, researchers hope to implicitly control for all inputs that affected test scores until that moment, thereby reducing the scope of bias. Nonetheless, there is no guarantee that omitted variables during the sixth grade, or earlier ones not captured by error-ridden pre-tests, do not continue to bias estimates.

Propensity Score Methods. The propensity score is a predicted probability that a student (or another unit of analysis) receives a treatment, given observed variables (Rosenbaum & Rubin, 1983). Using the previous example, researchers could estimate probabilities, with a probit or logit regression, that students attend a private school, conditional on family income ($X$) and other variables thought to influence either student outcomes or the likelihood of attending in a private school. There are numerous methods of incorporating propensity scores in the estimation of treatment effects (Imbens & Wooldridge, 2009; Stuart, 2010; Murnane & Willett, 2011; for examples in education policy, see Behrman, Cheng, & Todd, 2004; Hong & Raudenbush, 2006; Lara, Mizala, & Repetto, 2011; McEwan et al., forthcoming).

One method applies stratification (or subclassification) to divide observations into strata defined by similar values of the propensity score. A second method matches treatment observations to their nearest neighbor(s) in the comparison group—using the simple difference in propensity scores or another distance metric—and then discards unmatched observations. A third method uses the propensity scores to construct inverse probability weights that are used to re-weight treatment and control observations.

To estimate treatment effects, these methods can be used in concert with regression analysis (Imbens & Wooldridge, 2009; Stuart, 2010). Indeed, the use of inverse probability weights in a standard regression analysis yields estimates of treatment effects that are "doubly-robust." That is, they are consistent as long as either model—the logit or the linear regression—is correctly specified (Hirano & Imbens, 2002; Imbens & Wooldridge, 2009).

Propensity score methods do not impose arbitrary assumptions of linearity on the relationship between outcomes, policy variables, and other controls (Ravallion, 2005). This is particularly important when treatment and control observations are observationally quite different, and linear regressions would rely too heavily on extrapolation (Stuart, 2010). Propensity score methods also nudge researchers to explicitly analyze whether observed variables are well balanced across nonexperimental treatment and comparison groups. Yet, like regression analysis, the causal interpretation of evidence ultimately rests on the hard-to-verify assumption that unobserved variables are not correlated with outcomes and with the probability of receiving a treatment.

Randomized Assignment

In the 1990s, researchers in education policy grew disenchanted with the ability of statistical controls for observed variables to eliminate selection bias in nonexperimental data. A turning point was the widespread analysis and debate of results from a randomized experiment in Tennessee that identified the causal effect of smaller class sizes on student test scores. At the time, Krueger (1999) opined that "one well-designed experiment should trump a phalanx of poorly controlled, imprecise observational studies based on uncertain statistical specification" (p. 528). His opinion reflected a broader movement in social sciences to focus less on the rote application of statistical tools, and more on the quality of counterfactual reasoning deployed by researchers (Angrist & Krueger, 1999; Glewwe & Kremer, 2006). The statistician David Freedman (1991) remarked that "regression models make it all too easy to substitute technique for work" (p. 300) and called on researchers to expend more "shoe leather" in the pursuit of convincing causal arguments.

In the simplest randomized experiment, researchers flip a coin to determine which students are treated and which are not (noting that classrooms, schools, or districts can also be randomly assigned in cluster-randomized experiments). Thus, each student’s probability of receiving the treatment is an identical 0.5. The virtue of this approach is that it balances, by design, the distribution of students’ observed and unobserved characteristics across treatment and control groups. The two groups are not identical, of course, but they should be similar in large enough samples. Because control group members are similar, they provide an ideal counterfactual estimate of outcomes. To obtain an internally valid estimate of the causal effect, one estimates the mean difference between the outcomes of treated and untreated units (a pleasant irony of experiments is that credible evidence can be obtained with less sophisticated statistical methods). One could further apply statistical controls for observed differences between groups. If randomization proceeded without a hitch, then
doing so is not strictly necessary to eliminate selection bias, although it may improve the precision of estimates.

At the time this chapter appeared in the first edition of the Handbook, some lamented the dearth of randomized experiments in education (Cook, 2002). However, the last decade has been productive for methodological and applied research. There are many guides to the design, implementation, and analysis of randomized experiments in education and social policy (Orr, 1999; Shadish et al., 2002; Murnane & Willett, 2011; Glennerster & Takavarasha, 2013). There are primers on specific issues, including power analysis for cluster-randomized experiments (e.g., Dong & Maynard, 2013) and best-practice methods for the reporting of experimental results (e.g., What Works Clearinghouse, 2013). Experiments routinely appear in journals such as Educational Evaluation and Policy Analysis, the Journal of Policy Analysis and Management, and the Journal of Research on Educational Effectiveness. For example, experiments have been used to explore the effects of teacher performance incentives (Springer et al., 2010), whole-school reform (Borman et al., 2005; Cook et al., 2000), and private school vouchers in Washington, DC (Wolf et al., 2013), and New York City (Howell & Peterson, 2002; Krueger & Zhu, 2004), among many other areas of study.

In the voucher experiments, for example, researchers randomly awarded private school tuition vouchers to some students, so the treatment consisted of a “voucher offer.” In practice, not all students offered a voucher actually used it to enroll in a private school; by the same token, some students randomly denied a voucher still enrolled in a private school. This highlights a common feature of almost all social experiments: a subset of randomly assigned participants do not comply with the initial assignment.

This is not a flaw of the research design. One option is simply to compare the outcomes (test scores, in this case) of the full treatment and control groups as initially assigned, regardless of whether or not they take up the voucher offer. This yields an unbiased estimate of the aptly named “intent-to-treat.” Though not an estimate of the effect of actually using a voucher to attend a private school, it provides valuable information to policymakers. To further recover the effect of actually enrolling in a private school—among students induced to do so by the offer—researchers can use instrumental variables methods discussed in a later section.

Randomized experiments are not without pitfalls (for varied opinions, see Burtless, 1995; Heckman, 1995; Shadish et al., 2002). A common critique is that experiments, especially small-scale ones, yield causal evidence with less external validity than do quasi-experimental evaluations of scaled-up treatments in representative samples. As one example, most experiments begin with a nonrandom, convenience sample of sites, schools, or students (McEwan, 2013; Olsen et al., 2013). The subjects are often chosen for their enthusiasm, willingness to be randomized, or physical proximity to implementers or evaluators. (In the voucher experiments, for example, members of the experimental sample voluntarily applied to receive a tuition voucher.) Assuming that treatment effects are not homogenous, the extrapolation of experimental results to a well-defined, policy-relevant population is not straightforward. It requires the subjective application of rules of thumb (Shadish et al., 2002) or the use of statistical methods based on potentially strong assumptions (Hotz, Imbens, & Mortimer, 2005).

Experimental researchers with large enough samples often assess heterogeneity of results by estimating causal effects within subsamples of students or schools, dividing them by location, income, race, or other variables. However, subgroup analysis can resemble a “hunt” for any statistically significant effect within ever-smaller subsamples of data (Deaton, 2010). For this reason, some researchers tie their hands by prospectively registering analysis plans (see, for example, http://www.ridie.org and https://www.socialscienceregistry.org).

**Discontinuity Assignment**

One of the most credible quasi-experimental methods is the regression-discontinuity design (RDD) (Hahn, Todd, & van der Klaauw, 2001; Cook, 2008; Lee & Lemieux, 2010). In the most common RDD, students, schools, or other units are assigned to a treatment group on the basis of a single assignment variable—often achievement or income, but potentially any continuous variable—and a prespecified cutoff value. Reardon and Robinson (2012) consider extensions in which assignment is determined by more than one variable and cutoff (for instance, multiple test scores). To provide a simple illustration, suppose that a thousand students vie for college financial aid by taking a pre-test (the assignment variable). Students with scores of 50 or above (the cutoff) receive aid, and those with scores below 50 do not. As long as there is some error in the measurement of...
the assignment variable—quite likely in the case of a test score—a student’s precise location on either side of
the cutoff is credibly random (Lee & Lemieux, 2010). Thus, students close to the cutoff should be similar, on
average, not just in their values of the pretest, but in other observed and unobserved variables that determine
outcomes. The causal effect is estimated by comparing the outcomes of treatment and control students whose
values of the assignment variables are close to 50, using nonparametric or parametric methods that are the
subject of specialized literatures. Lee and Lemieux (2010) emphasize the inherently visual nature of
identification in the RDD and its vital role as a complement to model estimation.

Among a growing number of papers in education policy, the RDD has been applied to estimate the effects of
bilingual education (Chin, Daysal, & Imberman, 2013), college financial aid (van der Klaauw, 2002), early
childhood education (Gormley & Gayer, 2005; Ludwig & Miller, 2007), teacher training (Jacob & Lefgren, 2004),
and compensatory education for disadvantaged children (Chay, McEwan, & Urquiola, 2005). A hallmark of the
RDD in education is that researchers usually do not specify cutoffs or implement the assignment. Instead,
researchers take advantage of cutoff-based assignment that administrators used to allocate resources in a
transparent, fair, or efficient way (e.g., needy or meritorious students receive financial aid, low-scoring schools
receive assistance or sanctions and high-scoring ones receive rewards, less-effective teachers receive training,
and so on). The unintended usefulness of such rules to researchers has only recently been noted in many cases,
even when discontinuity assignment has a long history, as in the Head Start program (Ludwig & Miller, 2007).

What are the potential pitfalls of using discontinuity-based assignment? The most serious, related to internal
validity, is that students, or others subject to discontinuity assignment, are familiar with the treatment, the
assignment variable, and the value of the cutoff. If they have incentives to receive the treatment, or not, then
they may well attempt to manipulate their values of the assignment variable, thus introducing selection bias
(Lee & Lemieux, 2010; McCrary, 2008). Precise manipulation of many continuous assignment variables is
actually harder than it might seem (Lee & Lemieux, 2010). While most families can probably influence their
child’s pre-test score, for example, random errors in testing make it unlikely that they can affect it within a
narrow band of scores around the assignment cutoff. Students within this band contribute the most to
estimates of causal effects, implying that assignment variable manipulation must be very precise to bias
regression-discontinuity effects.

To test for manipulation, researchers typically search for suspicious clustering of students on either side the
cutoff (McCrary, 2008). They also compare students’ observed characteristics near the cutoff, which should vary smoothly across the break, in the absence of manipulation.

A frequent caveat of the RDD is that it is a “local” randomized experiment. Thus, its estimates are most
generalizable to subjects with values of the assignment variable near the cutoff, which could limit external
validity. For several reasons, this argument is sometimes overstated. First, “local” RDD estimates are often
obtained from large, administrative datasets with statewide or national coverage. In contrast, many
experiments report average treatment effects in less representative convenience samples. It is difficult to make
a priori judgments about which estimate has greater external validity, and it depends on the specific policy
question. Second, Lee and Lemieux (2010) note that RDD estimates are a weighted average treatment effect
across the entire sample, with weights proportional to the likelihood of drawing a value of the assignment
variable close to the cutoff. Ironically, an imprecisely measured assignment variable—say, a very noisy test
score—would imply that an RDD estimate is applicable to a larger proportion of the sample. Third, the cutoff is
sometimes of direct policy interest, since policymakers may use results to increase or decrease treatment
eligibility (van der Klaauw, 2002). Sometimes policymakers are interested only in effects on students far from a
cutoff, including the very poorest or low-achieving students. In such cases, the standard RDD is less credible,
and researchers must invoke additional assumptions and methods (Angrist & Rokkanen, 2012).

**Instrumental Variables**

The assignment of students, schools, or other units to policy treatments is often neither random nor entirely
based on values of an observed assignment variable. Besides controlling for observed variables, what remaining
methods are available to identify the causal effect of policies on outcomes? One of most popular in the last two
decades has been instrumental variables (IV) (Wooldridge, 2010; Angrist & Pischke, 2009).

In nonexperimental data, the receipt of policy treatments is usually correlated with unobserved characteristics of individuals that affect outcomes. Therein lies the empirical dilemma. Yet, some individuals in
the sample might receive a treatment because of luck or because they were encouraged to do so for reasons unrelated to outcomes. The challenge is to base estimates of causal effects entirely on "clean" variation in treatment status—that is, variation uncorrelated with unobserved characteristics that affect outcomes. It is easier said than done.

One must identify an instrumental variable, or instrument, that fulfills two conditions (Bound, Jaeger, & Baker, 1995; Wooldridge, 2010). First, it must be strongly correlated with the probability of receiving an intervention. This condition, straightforward to test in the data, is needed to ensure that the instrument actually induces students to alter their treatment status. Second, the instrument cannot be correlated with unexplained variation in the outcome variable (that is, the variation in outcomes that remains after controlling for other independent variables). The validity of the second assumption, more complicated to empirically test, usually rests upon the compelling reasoning of the researcher.

In applications to education policy, instruments are often related to features of geography or students’ location, which are assumed to be random in some regard, and thus viable candidates to fulfill the second condition. Towards estimating private school effects, Figlio and Ludwig (2012) show that the availability of subway transportation in metropolitan areas affects the probability that families, especially poorer ones, choose private schools. Using this as an instrumental variable, their analysis suggests that private school attendance has strong effects on reducing some risky teenage behaviors. Their analysis must assume that transportation availability is uncorrelated with student outcomes, controlling for other variables such as family income.

Hoxby (2000) estimates the effects of competition among public school districts on students’ outcomes such as test scores (see Rothstein, 2007, for a reanalysis). She measures competition as the concentration of public school districts within metropolitan areas; areas dominated by a few districts are assumed to be less competitive. The measure of competition is likely correlated with unobserved features of local students, schools, and communities that affect test scores. Hoxby argues that the number of streams in metropolitan areas (the instrumental variable) increases competition, because higher transportation costs led many areas to fragment into smaller school districts. The IV results suggest that metropolitan competition (induced by streams) has strong effects on test scores, based on the assumption that local geography is not correlated with unexplained test scores.

In each example, the validity of the second condition is hard to prove, and counterexamples are easy to invent. Do metropolitan areas with extensive subways have progressive mayors that invest in public schools? Do metropolitan areas with many streams and districts also have greater segregation by race or socioeconomic status that lowers achievement? In the most convincing IV analyses, there are a priori reasons to believe that instruments are uncorrelated with unexplained outcomes. This usually happens because institutional rules mandate the use of lotteries to allocate resources. In U.S. states, charter schools are legally obligated to use admission lotteries when the number of applicants exceeds available space (McEwan & Olsen, 2010). Abdulkadiroglu et al. (2011) use the results of Massachusetts lotteries as a credible instrumental variable for their treatment variable (the number of years enrolled in a charter school). Among students subject to specific school-by-year lotteries, the lottery outcome is plausibly uncorrelated with their eventual outcomes (except through its effects on charter school enrollment). Thus, IV estimates of the effects of charter school enrollment on student outcomes have strong internal validity. They are still subject to caveats about external validity. The estimates can be generalized to charter schools with sufficient demand to use lotteries, which are not necessarily representative of all charter schools (McEwan & Olsen, 2010). More generally, IV estimates can be generalized to students induced to take up a treatment due to variation in the instrument (e.g., students induced to enroll by virtue of winning a lottery). In the example, and IV research generally, these students are not always the population of interest (Deaton, 2010).

**Difference-in-Differences**

Difference-in-differences (DD) methods attempt to control for unobserved variables that bias estimates of causal effects, aided by longitudinal data collected from students, schools, districts, or states. Researchers employ two varieties of longitudinal data. Panel data track the progress of the same students or teachers in successive months or years. Repeated cross-section data follow different groups of individuals (e.g., second graders in successive years) that are clustered within the same schools, districts, or states.
The logic of DD causal inference is best communicated with an example based on repeated cross-section data (for its empirical implementation, see Dee & Levine, 2004). Of two states, Massachusetts and Maine, suppose that the former implements a school finance reform—increasing state financing of local public school districts—and the latter does not. To estimate the reform’s impact on a district-level variable (such as state revenue per pupil), a naïve approach would simply compare revenues between states in a single year of post-reform data. The difference is likely biased by selection, since unobserved variables could explain all or part of it.

Now consider the same difference, but in an earlier, pre-reform year of data. Evidently the difference cannot be attributed to a Massachusetts reform that has yet to occur. Pre-reform differences are due to unobserved variables between states that contaminated the naïve estimate. To control for these unobserved variables, the DD estimate of the reform’s effect subtracts the second difference from the first. The remaining “difference-in-differences” could be plausibly attributed to the reform. For this conclusion to be credible, the change in Maine’s per-pupil state revenues must be a good counterfactual for the change in Massachusetts’ per-pupil state revenues. Yet, suppose that state revenues per pupil rose more quickly in Massachusetts than in other states—even before the reform—because of economic growth due to a strong biotech industry. The DD will nonetheless attribute faster outcome growth in the treated state to the causal effect of reform.

In light of this pitfall, one of the best ways to assess the internal validity of DD results is to compare the trends of outcome variables across treatment and control groups before application of the treatment (Angrist & Krueger, 1999; Meyer, 1995). Evidence of similar trends bolsters confidence in the DD assumption. Dee and Levine (2004) estimated the effect of Massachusetts’s state finance reform on districts’ per-pupil state revenues. As controls, they used districts in Maine and Connecticut, which did not apply reforms, over the same pre- and post-reform period. DD estimates showed significant effects of the reform on per-pupil revenues. To support the use of these comparison groups, they showed that the variable had similar trends in the three states in years prior to the reform.

There are many variants of DD analyses, depending on the context, research question, and data. Dynarski (2003) leverages the fact that the Social Security Administration used to provide large college subsidies to children with deceased parents, but abruptly stopped doing so, beginning with the high school class of 1982. Dynarski identifies students with deceased parents—the treatment group—who graduated from high school before and after the change, and students without deceased parents—a comparison group never eligible for the benefits—over the same period. The DD estimates suggest large effects of subsidies on college attendance. In the United States, it is common to use repeated cross-section data on all 50 states, during periods in which some states were exposed to reform (perhaps at different times), and other states were not. The strategy has been applied, for example, to the effects of state accountability reforms (Hanushek & Raymond, 2005) and state finance reforms (Murray, Evans, & Schwab, 1998).

Researchers also apply DD methods to student-level panel data on test scores, guided by a similar logic of causal inference. Some students’ outcomes are observed before and after exposure to a treatment. Their outcomes are compared to students never exposed to the treatment. In Rouse’s (1998) seminal reanalysis of data from the Milwaukee voucher program, the treatment group consists of students observed before and after their selection to receive a private school voucher. The comparison groups consist of (1) students who were denied a private school voucher and (2) students in public schools who never applied for one. The DD estimates suggest that treated students have faster gains in math scores, but not in reading, than students in both comparison groups. The necessary assumption, as in previous analyses, is that treated students would have had trends in achievement similar to untreated students in the absence of the treatment. The assumption is not always tenable in student panel data. Suppose, for example, that students are more likely to apply to charter schools after experiencing slower test score growth in public schools than other students. If they switch to charter schools, a DD estimate of charter school effects could mistake this preexisting trend for part of a charter school “effect.” With more than two years of data, researchers can implement more complex models that compare changes in test score gains rather than levels (Bifulco & Ladd, 2006; Hanushek et al., 2007; Sass, 2006). Repeated cross-section studies have employed a similar strategy, controlling for unit-specific linear time trends (see the state-level studies of Loeb & Page, 2000; Hanushek & Raymond, 2005).

Combining Methods to Improve Causal Inference

96
Researchers usually apply more than one method in the same study to confirm that causal results are robust. Nearly all studies apply statistical controls for family and student characteristics that affect outcomes. In experiments and the RDD, controls are not strictly necessary since randomized or "local" randomized assignment credibly removes selection bias (although controls should not dramatically alter conclusions, implying a handy robustness check). An exception to this rule is when nonrandom attrition from treatment and/or control groups potentially creates imbalance in the distribution of observed or unobserved variables that affect outcomes. Besides regression-based controls or the use of propensity score methods, experimental researchers often attempt to construct plausible bounds for average treatment effects (e.g., Lee, 2009). When using DD strategies, researchers should always assess robustness to the inclusion of time-varying control variables, if such data are available. Finally, IV strategies are far more credible if the judicious use of controls helps rule out potential sources of correlation between the instrument and unexplained outcomes.

Researchers often combine DD methods with experiments (Krueger & Zhu, 2004; Skoufias, 2005), the RDD (Chay et al., 2005; Jacob & Lefgren, 2004), and IV strategies (Kuziemko, 2006; Loeb & Page, 2000). In experiments and the RDD, there should be no pretreatment difference in outcomes across treatment and control groups, so using longitudinal data is not strictly necessary to control for selection bias. But again, it provides a useful check of internal validity, and might improve the internal validity of research with nonrandom attrition.

Finally, researchers combine IV methods with randomized experiments and the RDD, in order to address imperfect compliance of students with policy treatments. In the New York City voucher experiment, for example, students were randomly assigned to receive a voucher, but not all students accepted the offer and actually enrolled in a private school. To recover the effect of actually enrolling—among students induced to do so by the voucher offer—researchers used the offer as an instrument for private school attendance (Howard & Peterson, 2002; Krueger & Zhu, 2004). The instrument plausibly fulfills both conditions: (1) it is correlated with private school attendance, and (2) the initial random assignment of the offer ensures that it is not correlated with unexplained outcomes.

In the analogous RDD case, units do not always comply with their initial cutoff-based assignment to treatment or control groups. RDDs with perfect compliance are called “sharp” designs, and those with imperfect compliance are "fuzzy" (van der Klaauw, 2002; Shadish et al., 2002). In the earlier example, families whose child is ineligible for financial aid (by virtue of obtaining a pretest score below 50) may nonetheless lobby to have the decision overturned. Other families may turn down the assigned treatment. In this fuzzy RDD, the effect of actually receiving financial aid—among students induced to do so by cutoff-based assignment—can be recovered by instrumenting the receipt of financial aid with the initial treatment assignment (van der Klaauw, 2002).

Conclusions

This chapter has described quantitative research methods used to estimate the causal effect of policies on education outcomes. Some, such as regression and propensity score methods applied to nonexperimental data, are ubiquitous but not always capable of delivering strong causal conclusions. Others, such as experimental and discontinuity research designs, are increasingly common, especially in the last decade. Indeed, there is now wide consensus among interdisciplinary education policy researchers about the need to deliver results with high internal validity.

Good causal research is a necessary but not sufficient condition for designing and implementing good policy. Notwithstanding this chapter’s emphasis on rigorous causal research methods, it does not address methods for answering descriptive or process questions nor does it review qualitative research methods (King, Keohane, & Verba, 1994). Both can be eminently scientific (Shavelson & Towne, 2002) and deserve serious attention from newer generations of researchers in education finance and policy. Finally, the growing number of high-quality studies has refocused attention on the inherent challenges to generalizing results to new settings, to diverse populations of students, to variations in the treatment, and to a wider set of outcomes. By the third edition of this Handbook, one hopes that the research pendulum will shift even further to issues of external validity and effective use of research in policymaking.
Notes


2. National samples, notably NELS:88 and ECLS, have been collected by the National Center for Education Statistics (NCES) and are available from their Web site (http://www.nces.ed.gov).

3. These include data from Chicago Public Schools (Jacob & Lefgren, 2004), Texas (Hanushek et al., 2007), North Carolina (Bifulco & Ladd, 2006), Florida (Sass, 2006), and teachers in New York State (Lankford, Loeb, & Wyckoff, 2002). The National Center for Analysis of Longitudinal Data in Education Research (CALDER) collects and analyzes administrative panel data from many states (see www.caldercenter.org).

4. Shadish et al. (2002) describe the history of the counterfactual reasoning. It has been formalized by statisticians, especially Donald Rubin (Holland, 1986), in a framework that has been widely adopted by econometricians (Wooldridge, 2010; Angrist & Pischke, 2009) and applied in research on education finance and policy.

5. For reviews of the early nonexperimental research, see McEwan (2000), Ladd (2002), and Neal (2002).

6. In this chapter, regression analysis implies ordinary least-squares regression (OLS). In education policy, it is common to apply multilevel or hierarchical models (Raudenbush & Bryk, 2002; Somers, McEwan, & Willms, 2004) that model error components and account for the potential correlation of errors within classrooms, schools, communities, or states. In so doing, they avoid understating standard errors of coefficients and overstating their statistical significance. Economists are more likely to report OLS coefficient estimates accompanied by adjusted Huber-White standard errors that allow for arbitrary correlations among units within clusters (Wooldridge, 2010). In comparisons, OLS with adjusted standard errors and other multilevel models yield similar results, though OLS with standard errors not adjusted for clustering can dramatically underestimate standard errors (Angeles & Mroz, 2001). This issue is not discussed further, but the research cited in this chapter generally reports cluster-adjusted standard errors.

7. This assumes a binary policy intervention (treated or not), though the discussion can be generalized to continuously measured policy interventions (e.g., class sizes of one to 50).

8. It is common for researchers to conjecture about the direction of bias, in concert with the oft reasonable presumption that selection on unobserved characteristics, such as motivation, might work in the same direction as selection on observed characteristics, such as family income (e.g., Somers et al., 2004). For a rigorous application of this reasoning to the effects of private school attendance on test scores, see Altonji, Elder, & Taber (2005).


11. Several studies found that nonexperimental statistical approaches, including regression and propensity score matching, do a poor job of replicating experimental results (Agodini & Dynarski, 2004; Glewwe et al., 2004; Glazerman, Levy, & Myers, 2003).

12. See Mosteller (1995) and Krueger (1999). In developing-country research, a similar role was played by the large-scale experimental evaluation of PROGRESA, a Mexican program that awarded cash payments to families in exchange for participating in health and education programs (Skoufias, 2005).

13. A coin flip or similar mechanism is only the simplest approach to designing randomized assignment. The essential point is that students or other units have well-defined probabilities of being assigned to the treatment.

14. In terms of the previous regression framework, randomized assignment ensures that the treatment (e.g., P) is uncorrelated with the error term.

15. Lee and Lemieux (2010) provide a detailed review of methods used to estimate effects in the RDD. In smaller samples near the cutoff, researchers typically use local linear regression; in larger samples that include students further from the cutoff, research flexibly model the relationship between outcomes and the assignment variable with higher-order polynomials. Lee and Lemieux (2010) discuss methods for the selection of sample bandwidths and polynomials.

16. Such manipulation is perhaps more likely when using discrete assignment variables that individuals or organizations—with incentives to obtain or avoid treatment—can precisely set. For example, many researchers note that day of birth is an assignment variable, and that children born after a specified assignment cutoff date are subject to treatments, such as delaying school enrollment by one year (McEwan &
Of course, motivated parents can manipulate day of birth by cesarean section or induced labor. While such manipulation might seem unlikely, Dickert-Conlin and Chandra (1999) show that U.S. parents precisely time year-end births to obtain tax benefits.

References


Hirano, K., & Imbens, G. W. (2002). Estimation of causal effects using propensity score weighting: An
application to data on right heart catheterization. *Health Services & Outcomes Research Methodology, 2*, 259–278.


Introduction

Measurements of academic achievement have a long history, but their frequency, heterogeneity, and, most importantly, scale have changed. Commencing as a pilot project in the 1960s, the international administration of tests of academic achievement has proliferated. The first tests were attempted in 12 countries with nonrepresentative samples, and the most recent tests were applied in 65 countries. International Large-Scale Assessments (ILSAs), in the future, will likely be applied in different jurisdictions—states (perhaps in Brazil, Russia, the United States, and India) as well as municipalities. The expansion of ILSA has been driven by demand for results from educational and political leaders, academics, journalists, and those who forge public policy, including economic policy. This growth has led to an increase in visibility and debates over interpretations and the implications of the results. This chapter provides an overview of the expansion of the ILSA, the impact of ILSA results on education policies in various countries, and a summary of the debates. It offers, albeit briefly, some guidelines as to their future utility.

Background: The Expansion of ILSA

International large-scale assessments (ILSAs) have become more important as policy-makers utilize their results in different ways. The relatively recent emphasis on educational accountability and the movement towards evidence-based research, as well as outcomes-based and standards-based education to inform public policy, account for the appeal of ILSA (von Davier et al., 2012; Steiner-Khamsi, 2003). International testing generally consists of large-scale assessments and surveys administered in different countries that provide comparative information both within and between countries (Wiseman, 2010). The main justification is that education in one country might be best understood by comparing it to education in other countries. Cross-national studies provide unique opportunities to understand issues in education and provide diagnostic and decision-making information about how to improve students’ learning (Bradburn & Gilford, 1990; Ma, 1999; Stigler & Hiebert, 1999; Cai, 2004). Cross-national comparisons can also provide a chance to examine educational factors outside the local context and to explore which are effective in which cultural contexts (Reynolds, 2000).

The origins of ILSA can be traced to the late 1950s, when Benjamin Bloom, C. Arnold Anderson, and colleagues at the University of Chicago agreed that the world might be seen as a "one big educational laboratory" (Husen, 1967; Heyneman & Lykins, 2008). It was noted that although the majority of research on education came from the United States, the country accounted for only 3 percent of the world’s schools and school children. The idea was to see the differences in participating countries’ educational systems as natural variations and to take advantage of this international variability to determine which characteristics might be more closely associated with academic achievement and under which circumstances (Husen, 1973). The first pilot survey of an international comparative assessment in reading, science, and mathematics was administered to nonrepresentative samples of 9,918 13-year-old students in 12 countries (Postlewaite, 1975). This informal
group of academics founded a nongovernmental organization called the International Association for the Evaluation of Educational Achievement (IEA). The IEA raised enough money to sponsor a pilot study in 1964 called the First International Mathematics Study (FIMS). The style of organizing this first study continues a half century later. Curriculum and pedagogy specialists from different countries compare formal curricula, textbooks, and teacher training. Achievement test items are then developed out of these comparisons.

Since the FIMS study, the ILSAs have expanded in both reach and breadth (Rutkowski et al., 2013). FIMS was followed by the Second International Mathematics Study (SIMS) in 1980–1982, the Third International Mathematics and Science Study (TIMSS) in 1995 and its repeat in 1998. The IEA also administers the Progress in International Reading Literacy Study (PIRLS), which assesses reading skills of nine- and ten-year-olds and collects information on student and school environments. The field continues to expand and includes assessments and surveys covering topics such as computers in education, civics education, teacher education (Teacher Education and Development Study in Mathematics: TEDS-M; Teaching and Learning International Survey: TALIS), and adult literacy (Programme for the International Assessment of Adult Competencies: PIAAC) (Rutkowski et al., 2013).

Table 7.1 illustrates that the types of international testing were diversified over time, and the number of participating countries has increased, particularly after the 2000s. For instance, the number of countries participating in TIMSS increased from 45 in 1999 to 63 in 2011. One of the currently best-known assessment programs, the Program for International Student Assessment (PISA) was organized by the Organisation for Economic Co-operation and Development (OECD) and first administered to its member countries in 2000. Participation in this assessment of mathematics, science, and reading for 15-year-olds doubled over the next decade. While 32 countries participated in PISA 2000, 65 countries participated in 2012, the most recent assessment of PISA, because the OECD invited non-OECD member countries to participate (Kamens, 2013).

**Distinction Between TIMSS and PISA**

Because TIMSS and PISA are the most widely known international tests, we describe them in more detail. As Table 7.1 illustrates, TIMSS 1995, administered by the IEA, was the first assessment to bring together mathematics and science in a single study (Mullis & Martin, 2007). To reflect the fact that the demand for international test scores
had become permanent, TIMSS was renamed Trends in International Mathematics and Science Study (TIMSS) in 2003, and it assesses mathematics and science achievement of fourth and eighth graders every four years.

One main feature that distinguishes TIMSS from PISA (which also tests mathematics and science) is that TIMSS is a curricula-based test. All IEA-sponsored studies, including TIMSS, collect information on the international variations in curricula, including variations in goals, intentions, and sequences of curricula (Robitaille et al., 1993). TIMSS also collects information on what the pupil learned (attained curriculum), what the teacher is expected to teach (intended curriculum), and what the teacher has in fact taught (implemented curriculum). TIMSS results in mathematics and science reflect how much students have learned within the given curriculum in each country. The relevance to the national curriculum may help determine the level of local education policy relevance, particularly important in those countries with few other sources of comparable student information (Elley, 2002).

Because of the way in which items are developed, TIMSS has in depth information on curriculum and teaching practices. But TIMSS also includes results from separate questionnaires administered to teachers and students and an assessment of physical facilities in sampled schools. The information from TIMSS provides a rich source to explore the influences on academic achievement. However such depth requires considerably more time and cost than ILSA tests without such background information.

PISA is a relative newcomer in the field of international testing, as the OECD became involved in international education testing only in the late 1990s (Wiseman, 2010). PISA was developed in 1997 at the request of the OECD member states. Their concern was to assess the degree to which students were prepared to enter a "knowledge society" (Lundgren, 2011). PISA tests are designed by test developers contracted by the OECD on the basis of what they consider should be the normal performance level of someone at a given age regardless of a nation’s curriculum. Because items were not drawn from each nation’s curriculum, a principal virtue of PISA was the fact that it could be administered quickly and comparatively cheaply. PISA was specifically adapted to the demands of political leaders who required a snapshot of student performance more than they required information on the influences on that performance.

Hence, PISA and TIMSS are different from one another not only in content but also in construction and purpose. PISA tests 15-year-olds’ knowledge and skills (competencies) in reading, mathematics, and science.

<table>
<thead>
<tr>
<th>Sponsor</th>
<th>Description</th>
<th>Number of participating countries</th>
<th>Year(s) conducted</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEA</td>
<td>First International Mathematics Study (TIMSS)</td>
<td>13</td>
<td>1995</td>
</tr>
<tr>
<td>IEA</td>
<td>Six subjects study</td>
<td>10</td>
<td>1995-1996</td>
</tr>
<tr>
<td>IEA</td>
<td>Science</td>
<td>10</td>
<td>1995-1996</td>
</tr>
<tr>
<td>IEA</td>
<td>Reading</td>
<td>10</td>
<td>1995-1996</td>
</tr>
<tr>
<td>IEA</td>
<td>Language</td>
<td>10</td>
<td>1995-1996</td>
</tr>
<tr>
<td>IEA</td>
<td>Foreign language (foreign language)</td>
<td>10</td>
<td>1995-1996</td>
</tr>
<tr>
<td>IEA</td>
<td>Civic education</td>
<td>10</td>
<td>1995-1996</td>
</tr>
<tr>
<td>IEA</td>
<td>First International Science Study (ISIIS)</td>
<td>19</td>
<td>1997-1998</td>
</tr>
<tr>
<td>IEA</td>
<td>Second International Mathematics Study (SIMS)</td>
<td>10</td>
<td>1998</td>
</tr>
<tr>
<td>IEA</td>
<td>Second International Science Study (SISIIS)</td>
<td>10</td>
<td>1998</td>
</tr>
<tr>
<td>IEA</td>
<td>Third International Mathematics and Science Study (TIMSS)</td>
<td>10</td>
<td>2000</td>
</tr>
<tr>
<td>IEA</td>
<td>Third International Science Study</td>
<td>10</td>
<td>2000</td>
</tr>
<tr>
<td>OECD</td>
<td>Program for International Student Assessment (PISA)</td>
<td>69</td>
<td>2000</td>
</tr>
<tr>
<td>OECD</td>
<td>Language Education Study</td>
<td>50</td>
<td>2003</td>
</tr>
<tr>
<td>OECD</td>
<td>Civic Education Study</td>
<td>45</td>
<td>2003</td>
</tr>
<tr>
<td>OECD</td>
<td>Third International Mathematics and Science Study (TIMSS)</td>
<td>45</td>
<td>2003</td>
</tr>
<tr>
<td>OECD</td>
<td>Teacher Education and Development Study in Mathematics (TEDS-M)</td>
<td>35</td>
<td>2007-2008</td>
</tr>
<tr>
<td>OECD</td>
<td>Teacher and Learning International Survey (TALIS)</td>
<td>35</td>
<td>2007-2008</td>
</tr>
</tbody>
</table>

Source: (OECD) tables modified by author.
that are needed in the labor market, as opposed to mastery of school curriculum. PISA is independent of national curricula whereas TIMSS is based on an assessment of each nation’s curriculum. Barry McGaw (2002), the Director for Education of the OECD, described the difference this way: TIMSS is interested in "What science have you been taught and how much have you learned?" while PISA is interested in "What can you do with the science you have been taught?" (McGaw, 2002). PISA is also more policy oriented and provides recommendations in PISA reports, which makes it a strong tool for use in policymaking (Figazzolo, 2009).

Impact of ILSA on Education Policies

ILSAs now hold a unique place in the history of education. No previous innovation has captured the attention of a similarly broad range of interest groups, governments, unions, academics, and representatives of political organizations. Reactions to TIMSS and earlier IEA assessments were significant, but often transitory, given that they may take a decade to develop. With PISA however, the reaction can be immediate, and because PISA can be repeated every two or three years, political reactions can be prolonged. Following the announcement of PISA results there is time to alter policies and test the degree to which alterations may make a difference over time. Though low for the test-takers, the stakes are high for the organizations and institutions whose reputation the results may affect.

A questionnaire to policymakers in 65 countries found that PISA performance was nearly universally considered as having the “potential to define” policy problems and set the agenda for policy debate at both national and local levels (Breakspear, 2012). PISA results are said to have created an “inter-dependence” across European systems of education and now constitute the sine qua non of education policymaking in Europe (Grek, 2012). In some instances the political stakes are so high that scores have been withheld and have been a factor in challenges to governments in power, as was the case in Hungary (Bajomi et al., 2009). Because of the political reaction created by consistently poor results, the Republic of South Africa stopped participating in ILSA (Wiseman, 2013). Japan has interpreted scores very carefully so as to fit into a local interpretation and legitimize policies. The international discussion of PISA results, for instance, caused the Ministry of Education to question its yutori (low-pressure) curriculum and reestablish the ministry’s political centrality in a time of neoliberal state restructuring (Takayama, 2008).

The use of ILSA data to change education policy has increased partly due to the semantics of globalization (Schriewer, 2000), which has generated political and economic pressure to compare education systems across countries (Steiner-Khamsi, 2003). Whatever the reason, the tendency to use ILSA in national policymaking has exploded (DeBoer, 2010; Smith & Baker, 2001; Wiseman & Baker, 2005). With increased pressures for educational accountability and globalization, it is assumed that international testing will continue to expand (Kamens & McNeely, 2010). It is useful, therefore, to examine the impact of ILSA on national education policies. Table 7.2 summarizes the findings of recent studies that have examined this issue (e.g., Wagemaker, 2013; Heyneman & Lee, 2013). In this section, we focus on several useful themes and support each with illustrations from specific countries.

ILSA “Shock”

In some countries the results of the ILSA served as a wake-up call for the education sector. The shock value was strongest in countries that believed their education system had been of high quality. Traditionally the German education system had an underpinning philosophy that focused on individuals’ desire to develop within themselves (Sorkin, 1983; Neumann et al., 2010). Local states had authority over education governance; teachers had authority over student assessment (Fensham, 2009). There was no standardized testing.

The belief that German education system was superior to others was shaken by the unexpected results from TIMSS and PISA. The mediocre German performance in TIMSS 1995 was followed by an even worse performance in PISA 2000 (Lehmann, 2011; Neumann et al., 2010). The shock in Germany had an impact similar to that in the United States with the launch of Sputnik and the publication of the Nation at Risk report (Gruber, 2006). The mediocre test performance of German students led to several significant changes.
Policymakers increased the funding of schools, took measures to improve instructional quality, and introduced National Education Standards (NES) (Neumann et al., 2010). The PISA framework influenced the NES and related curricular reforms. German states also adopted a concept of German schools as self-managing organizations and strengthened their autonomy (Grek, 2009). ILSA results also shifted the academic debate, which had been didactic and normative, to one that is more empirical and practice focused (Bohl, 2004).

Table 7.2 Impact of ILSA in Selected Countries

<table>
<thead>
<tr>
<th>Countries</th>
<th>PISA</th>
<th>TIMSS</th>
<th>Changed curriculum reflecting ILSA test items</th>
<th>Established national education standards or standardized examination</th>
<th>Became aware of subgroups (immigrants, low-income)</th>
<th>Decentralized education finance and strengthened school autonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Chile</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Denmark</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>England</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Iran</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Japan</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Kyrgyzstan</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Latvia</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Macedonia</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Romania</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Russia</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

At first, France ignored its mediocre performance in PISA (Dobbins & Martens, 2012), but after the deterioration in performance between 2003 and 2006, the government attempted to institutionalize a results- and evaluation-based approach emphasizing international comparisons (Mons & Pons, 2009; Pons, 2011). PISA seems to have provided a point of convergence for center-right and center-left points of view, including the priority of transferring Finnish pedagogical methods and school autonomy policies (Dobbins & Martens, 2012).

Kyrgyzstan experienced similar consequences as a result of participating in PISA 2006. Kyrgyzstan took the last place among the 57 participating countries, and the poor results came as a shock (Shamatov & Sainazarov, 2010). The PISA results indicated a shortage of school resources and highlighted issues of equity in access. The results were disputed and precipitated a national sample-based assessment (NSBA) that validated PISA results (Wagemaker, 2013). Despite the dispute over the results, the government of Kyrgyzstan accelerated reforms in several areas, including development of new standards and curricula, a reduction of the number of subjects and education load per teacher, and the introduction of per-student financing (Shamatov & Sainazarov, 2010; Briller, 2009; Silova & Steiner-Khamsi, 2009).

**ILSA and Low-Performing Subpopulations**

In some instances ILSA results helped focus attention on specific low-scoring subpopulations. PISA and PIRLS results raised concerns about immigrant populations in Germany (Schwippert, 2007). PISA results revealed that students in Switzerland had shortcomings in reading competencies and significant gaps based on
socioeconomic background (OECD, 2002). On the basis of PISA 2000, Denmark implemented a range of reforms focusing on students who were disadvantaged socioeconomically and immigrants (Ege-lund, 2008).

**ILSA and the Support of Policies Already in Place**

In some instances the results were used to justify policies already in place and over which there had been controversy. Switzerland, for example, was already making efforts to harmonize the different educational standards in each canton. After the PISA results, this trend moved to the top of the political agenda (Bieber & Martens, 2011). In Ireland, PISA 2006 results speeded up changes already planned for the curriculum in lower secondary education (Figazzolo, 2009). In England, PIRLS 2001 revealed that their National Literacy Project (a large-scale pilot study of a new way of teaching literacy at the primary level) proved to be quite successful (Twist, 2007). Similarly, students in New Zealand showed high performance in PISA, reinforcing existing policies (Dobbins, 2010).

**ILSA Items and Curriculum**

In some instances the items on ILSA tests, which often emphasize higher order cognitive skills of assessment and evaluation, have served to influence those emphases in new curriculum. Educators in the Russian Federation analyzed the TIMSS and PIRLS frameworks and developed recommendations for new educational standards for primary schools (Kovaleva, 2011). Macedonia and Iran used TIMSS test items to develop their national test items (Elley, 2005; Kiamanesh & Kheirieh, 2000). Slovenia and Romania also reflected the content of TIMSS curriculum framework and assessment items in their national curricula (Klemencic, 2010; Noveanu & Noveanu, 2000). PISA 2003 has led Japan to revise its national curriculum to incorporate competencies tested in PISA (Breakspear, 2012).

In case of Chile, TIMSS 1999 prompted curricular reform, new content standards in 2009, and reform in teacher education (Cariola et al., 2011). Singapore used TIMSS as one of several sources to inform annual reviews of their national curriculum (Poon, 2012). Iran used TIMSS to develop its first national item bank for primary education and a national research project to develop new test items (Kiamanesh & Kheirieh, 2000).

**ILSA and National Assessments**

In some instances ILSA tests have stimulated countries to design and implement their own national assessments. This included the Czech Republic (Strakova et al., 2000) and Latvia, which established a centralized examination system for secondary education (Geske & Kangro, 2000). Russia did not have a tradition of using standardized assessment in schools, but began to use standardized tests after the TIMSS results were released (Kovalyova, 2000). Whether establishing national assessments will lead to better student performance is a separate question. Finland, for instance, one of the highest-achieving countries in PISA, does not implement a national assessment until the end of basic education (ninth grade) (Kupiainen et al., 2009).

**Countries Less Affected by ILSA Results**

Not every country is affected equally by the results of ILSA. If a country was already aware of its poor results, new below-average performance reiterates what had already been expected. For instance, the United States was not heavily affected by the ILSA results because Americans were already aware of their performance problems since the first Sputnik and A Nation at Risk (Wiseman, 2013; Bieber & Martens, 2011).

**Impact of Regional Assessments**

Perhaps the most significant reluctance to participate in ILSA has come from countries that do not expect to
perform well. Their resistance has been modified by the introduction of region-based ILSA in sub-Saharan Africa and Latin America. The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) began in 1995 with the agreement of 15 anglophone countries. Data on mathematics and reading literacy were gathered in 2000 and 2007. The Program d’analyse des systems educatifs des pays dela CONFEMEN (PASEC) was begun in 17 francophone countries with information gathered on French, mathematics, and occasionally a national language. UNESCO, UNICEF and the Inter-American Development Bank helped establish the Laboratorio latin americano de evaluation de la calidad de la educacion and sponsor tests to Monitor Learning Achievements (MLA) in the 1990s (Heyneman & Lee, 2013, p. 49).

Like PISA and TIMSS, regional assessments have had significant impact in some countries, whereas in other countries they were used to reinforce the status quo of current education policy. In Togo and Guinea teachers were being hired on contract with short periods of pedagogical training. PASEC results suggested that the teachers employed on a contractual basis were not doing any worse than teachers who are regularly employed, and current policies were continued (Bernard & Michaelowa, 2006).

As with global surveys, the impact of regional assessments tends to be larger when results are unexpected (Murimba, 2005). For instance, unexpected results led to policy change in Senegal where it was widely believed that a Senegalese primary student who repeats his/her grade would perform academically better than those who do not; however, the initial results of PASEC in 1998 showed no evidence to support that (Bernard & Michaelowa, 2006). After the continued negative result of grade repetition in 2003, the government prohibited grade repetition (Bernard & Michaelowa, 2006).

Perhaps the most important impact in Latin America has not been the ILSA tests directly but findings summarized, interpreted, and disseminated by an organization called the Partnership for Educational Revitalization in the Americas (PREAL). Since 1995, PREAL has published report cards that give grades to countries for their educational performance. Published in French, Spanish, Portuguese, and English, each report card has stimulated changes in education policy and practice. These have included new national education standards in Peru and Honduras and a guide for reform in standards in Chile. Recommendations from PREAL’s report cards have been directly implemented in El Salvador, Panama, and the Dominican Republic. PREAL’s demand for greater accountability has influenced changes in Jamaica, Brazil, and Mexico City; and although countries in Latin America have traditionally been reluctant to participate in global ILSA, PREAL’s recommendation to participate influenced countries to break with tradition and join PISA or TIMSS. The argument made by PREAL concerns the shift away from dictatorships toward democracy. PREAL views the results of ILSA as an illustration of the power voters have over the existence of information on which to judge their public school systems. Without these assessments, political leaders might claim excellence without reference to the facts; with assessments, claims of excellence can be compared to international standards of evidence (Heyneman & Lee, 2013, p. 50).

The Future of ILSA

Technical Debates

Though carefully designed, ILSA tests are not without controversy. In spite of the efforts put into their construction, some argue that they continue to lack validity or reliability (Prais, 2003; Bracey, 2004, 2009; Dohn, 2007). Others argue that test items should be constructed and used differently. The purpose of ILSA tests is abstract and distant from the test-taker. Instead, they suggest the purpose should focus on having students achieve concrete accomplishments, which certify important qualifications. Test-takers differ dramatically in their desire to perform well. Korean students believe their performance will reflect the prestige of their nation; American students are not similarly motivated (Baker, 2007). If tests could be restructured to certify concrete accomplishments, students and schools might regain much needed balance in the differences on the part of the test-taker to do well on them. Interpretations of test results usually concentrate on the differences between schools in accounting for differences in academic achievement. Some suggest that the more valuable approach might be to concentrate on the differences within rather than between schools. For instance the OECD PISA
report points out that some countries (Finland, Iceland, and Norway) have lower portions of variance explained by the differences between schools (considered an accomplishment); while in other countries (Hungary, Germany, Bulgaria, Austria, Belgium, Slovenia, and the Netherlands) the portion of variance explained by between-school differences is considerably greater. The distinction may matter. Policy recommendations generated by concerns over between-school variance might differ from those (such as to address immigrant differences) over within-school variance (Gaber et al., 2012).

The implications of research findings that use ILSA data are sometimes hampered because of problems in sampling, coverage, and administration (Wiseman, 2010). Sampling frames of ILSA tests are not the same across countries. Allowing for different sampling frames makes sense as each country has its unique context, but it also makes it difficult to compare the findings across countries. For example, sampling schools by location (rural, urban) might be important in countries like China where the achievement gap between rural and urban school students is large, but would not make much sense for countries like Singapore, where all schools are located in the urban area.

In addition to differing sampling frames, some countries may not fully adhere to the IEA or OECD’s regulations related to sampling representativeness. If a country samples schools limited to elite students it can present problems of international validity and may generate objections from countries whose samples adhere to the standards recommended in a report of the National Academy of Sciences (Bradburn & Gilford, 1990). PISA was applied in many regions of China, but only scores from Shanghai were reported publicly. In addition there have been questions about whether the *hukou system* in China has prevented a representative sample of 15-year-olds in the Shanghai student population from taking the test (Loveless, 2014). Thus, the high score of Shanghai in PISA 2009 is subject to criticism that it does not necessarily represent the student population in Shanghai.

Another limitation of ILSA is the absence of prior test scores. ILSA provides only cross-sectional data, making it difficult to draw causal inferences with these assessments. However, Carnoy (2006) suggests that PISA and TIMSS provide sufficient cross-sectional data to challenge important assumptions, and with their repeated applications, they can be used to approach causal inference on the effects of particular interventions.

The findings from ILSAs may also be hampered because of the cultural heterogeneity of participating countries. Can one generalize a major finding across countries with such diversity? Hanushek et al. (2013) address this issue through the use of country-fixed effects. Yet fixed effects may cause researchers to miss how country-specific characteristics affect student outcomes. This may be important for instance when comparing East Asian countries (where education is uniformly highly valued) with other parts of the world. The lesson is that drawing policy implications requires caution.

**Academic Debates**

Technical debates concern the nature of test design; academic debates concern the nature of cultural explanations, political effects, and institutional implications of the results. Some observers note the sudden dominance of the ILSA, organized by international organizations, and conclude that the nation-state has begun a withdrawal of sovereignty over education policy. Their concern is that

> the very meaning of public education is being recast from a project aimed at forming national citizens and nurturing social solidarity to a project driven by economic demands and labor market orientations.

(Meyer & Benavot, 2013, p. 10)

These concerns are underscored by some in the academic community who bemoan the dominance of neoliberalism and its association with demand-based reforms that favor school choice, privatization, centralized goals, and competition as ways to improve efficiency and effectiveness (Ravitch, 2011; Steiner-Khamsi, 2003; Daun, 2005). They worry that school systems may be reduced to bureaucratic objects subject to the controls of external auditors (Apple, 2005; Power, 1999). And they seem concerned that the rise of ILSA represents a power shift away from the nation-state toward international non-educational organizations.

This opinion has been common to those interested in the World Bank, but is new with respect to other international organizations. Traditionally most international organizations have been viewed as harmless. Some suggest that they are now able to pressure national governments to adopt new objectives and values.
(Finnemore, 1996; Barnett & Finnemore, 1999). Meyer and Benavot put it this way:

Like all bureaucracies, they (international organizations) make rules, and, in so doing, they create social classifications that frame and reframe our understanding of social practice. International Organizations define tasks, create new social actors, and transfer new models of political organization around the world. (2013, p. 12)

These new concerns about the dominance of international organizations may overestimate their degree of autonomy. OECD, for instance, has little room to maneuver outside of the directives and financing of its member states. Its priorities, methods, and reports directly reflect their interests. Projects such as PISA are financed by voluntary contributions from those states that wish to participate. And though there are significant differences in the level of financing by which OECD member states are assessed, each has an equivalent vote over OECD’s direction and products. Similarly, the World Bank does essentially what its member states demand. Although member states own quite different levels of equity shares, it is rarely the case that countries with more shares line up against those with fewer shares (Heyneman, 2003 and 2012).

Perhaps more compelling are the arguments that come from those who suggest that the position of dominance of ILSA represents a general shift toward what is seen as a rational audit explosion gaining ascendancy in all public organizations—public hospitals, utilities, social welfare, and a move toward comparing similar organizations across national borders. To them, this audit culture appears to have emerged in a general search for best practices and magic bullets (Kamens, 2013). The underlying rationale however has been the connection between the performance of education and other public institutions and economic growth, prosperity, and what is described as the knowledge economy. It seems simplistic to suggest that this connection exists, but some might go further and portray test scores in particular subjects such as mathematics as being able to influence economic growth (Hanushek & Woessmann, 2009). If the economic fate of the nation rests on the ease of entry to the knowledge economy, and if entrance is influenced by the spread of geometry knowledge, the importance of PISA results rises to a new level.

In contrast to these conceptions of the economic importance of TIMSS and PISA results some explain performance on the basis of unshakable local cultural tradition. These observations do not deny the importance of ILSA but rather focus on the cultural background to testing. Korean, Japanese, and Chinese cultures for instance, have depended upon examination performance for centuries. Examinations have represented one of the few means of open access to social mobility, and examination results have taken a place of singular importance within the family for acquiring status assurance of the mother. Tests have also been used not only to gain access to universities but to allocate students to particular institutions and programs of study (Tucker, 2011). Similarly, while it has been common to note that Finland’s success on PISA has come without the many demand-side reforms typical of North America—for example, teacher performance tests and frequent individual, school, state, and national assessments—what is less often noted is that Finland’s economy had virtually collapsed in the 1990s, and national education performance was widely believed to be necessary for national survival (Sahlberg, 2011). Such extreme consensus is not simple to replicate elsewhere.

Differing test cultures have also led to questions about what is meant by schooling. In many Asian countries families are so preoccupied with examination performance that they consider the public school only one of several necessary mechanisms to augment performance. About 8 in 10 South Korean high school students utilize private tutors in addition to the time spent in public schooling. Universal private tutoring may double the amount of time devoted to test preparation. This suggests that Korean test performance may be high but also inefficient. When time out of school preparing for tests is combined with the time spent in school in both South Korea and the United States, it turns out American pupils are more efficient than pupils in South Korea. PISA performance of South Korean students come at about 30 percent greater time cost than the typical student in North America (Heyneman, 2013). This result underscores the fact that schools and education more generally comprise only a portion of the explanation for differences in ILSA results. Particularly important may be the result of differences in poverty and the existence of social safety nets associated with lowering the effects of poverty (Ladd, 2012); although the impact of poverty may differ from one country to the next (Heyneman & Loxley, 1983; Gamaron & Long, 2007), perhaps because of differences in income inequality (Chudgar & Luschei, 2009).

Official interpretations of PISA results have led to many new and constructive ideas well aside from the traditional Olympic race to the top. For instance, new interpretations have emphasized international
comparisons of achievement gaps based on social background or performances changes over time sometimes associated with shifts in education policies (OECD, 2011; Schleicher, 2009).

Debates therefore have raised important issues but have not led to calls for the elimination of ILSA. The question then becomes what should we do with ILSA results? Is there an appropriate and constructive manner to use them?

Conclusions

ILSA performance will remain an important component of education in the foreseeable future. The question is how the education community should treat the results. Should it rail against the tests and their visibility in political debates (Meyer & Benevot, 2013, p. 14; Meyer, 2013)? Should the results be used to reinforce education policy positions held ex ante? Should a case be advanced for communities not to participate?

Systems of large-scale testing concentrate on particular content areas that are required and measurable. These can include knowledge that is expected to be used on a frequent basis, such as economic principles, scientific evidence, and the skills of synthesis and evaluation. They can also include knowledge of information that is not expected to be used on a frequent basis, such as the periodic table of elements (Feuer, 2012, p. 11).

But schools are expected to accomplish many goals other than in particular content areas. These may include the incorporation of characteristics such as diligence, empathy, social responsibility, and the normalcy of performing manual labor. These may include skills of leadership, cultural awareness, and the ability to care for animals. They may include proficiency in a second language or actions thought to foster particular outcomes such as community service and cross-cultural experiences. One problem of ILSAs is that they are not able to reflect the degree to which school systems accomplish these other goals well (Heyneman, 2005).

In addition to narrow coverage, ILSA results are associated with other weaknesses. On the basis of ILSA results, it is common to infer trends from snapshots in time, which may lead to generalizations that are premature. For example, in the 1980s many assumed that Japan’s achievement scores were responsible for its superior economic performance prior to the stagnation in economic performance in the 1990s. ILSA may be biased in other ways. For instance, there may be a tendency for scores to be biased upwards in countries experiencing population declines (Feuer, 2012, p. 11).

Attributing changes in economic performance to scores in mathematics achievement (Hanushek & Woessmann, 2009) may be tenuous given that economic performance depends on a wide variety of influences. Current students would not have an impact on labor market productivity without a time lag of a decade or more, but the association with economic performance rarely accounts for this. Even in high-performing countries, there are persisting internal variations such as gender gaps in Finland and high-performing states within the United States. Moreover, it may be the case that the direction of the influence is the opposite from what is assumed. Economic performance, for example, may have an influence on school performance (Feuer, 2012, pp. 17–18).

Regardless of economic performance, it may be dangerous to rely on ILSA results to determine education policy. When applied properly, international comparisons are used to inform; when applied improperly they are used to mimic. For instance, unlike students in the United States, students in Finland are rarely subjected to quantitative evaluations in their schools, yet they perform well on PISA (Sahlberg, 2011). But Finnish children from both rich and poor families have similar values with respect to education. Teachers may not have to face the same problems of classroom discipline as do teachers in the United States. Raising the stakes for American pupils is one method of instilling a desire to perform well among students who have problems of understanding the importance of their education. Where that importance is already well understood and already present across socioeconomic strata, frequent performance tests may not be necessary. The key is to not use the presence or absence of a sector policy in a high-performing country to dictate a transfer of that policy elsewhere. For instance, when referring to the use and interpretation of National Assessment of Educational Progress (NAEP) scores Feuer says:

The basic idea that the program can promote dialogue, rather than issue summative comparative judgments of quality of teaching or schooling in various locales, remains one of its distinguishing characteristics. By analogy, then, rather than view the results of ILSA programs as prima facie evidence of comparative success or failure, and by extension as the clinching argument for reforms that imitate characteristics
of schools systems where students perform better, it might make more sense to explore how different types of assessments reflect values and expectations of schooling and to use the results as catalyst for public conversation and debate.

(2012, p. 20)

One illustration of how tenuous it may be to extrapolate from one successful environment to another is the prevalence of shadow education among the Asian nations so successful in PISA. While it is true that PISA scores are high among pupils in South Korea, it is also the case that the typical Korean pupil spends the entire period of adolescence preparing for their tests, with little room for other development experiences or goals. Scores may be higher in Korea, but Korean adolescents have little experience locating employment or participating in sporting events, activities that may be as common an expected experience among American adolescents as participating in shadow education is for adolescents in Asia. In predicting future economic performance, which is more relevant? In determining balance among well-adjusted adults, which is more important? These are the questions that ILSA raise but cannot answer.

Notes

1. Until recently, state-level performance on the National Assessment of Educational Progress, the national assessment within the United States, was resisted by the southeastern states which anticipated their low performance by comparison to states in other regions.

2. The *hukou* system in China limits a family’s permanent residence to their place of birth. Work in urban areas must be on the basis of a temporary migrant. Children of migrants are not allowed to attend regular public schools outside their place of birth. This significantly limits the representativeness of children in public schools in urban areas. The *hukou* system in China is parallel to the systems in North Korea, Vietnam, and the former Soviet Union.

References


made at Seventh Central Asian Forum on Education Organized by UNICEF (September 15–17), Bishkek, Kyrgyzstan.


Heyneman, S., & Lee, B. (2013). The impact of international studies of academic achievement on policy and research. In D. Rutkowski, L. Rutkowski, & M. von Davier (Eds.), Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis (pp. 37–72). Boca Raton, FL: CRC.


Schreiner, J. (2000). World system and interrelationship networks: The internationalization of education and


Wagemaker, H. (2013). International large-scale assessments: From research to policy. In D. Rutkowski, L. Rutkowski, & M. von Davier (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 11–36). Boca Raton, FL: CRC.


Section II
Making Money Matter

SECTION EDITOR: JENNIFER KING RICE
Toward an Understanding of Productivity in Education

JENNIFER KING RICE AND AMY ELLEN SCHWARTZ

The growing national interest in the academic performance of public school students has heightened interest in understanding how schools, teachers, and other educational resources can be marshaled to improve student performance. Because the public demand for improved public school outcomes is not matched by a similar willingness to increase spending, there is considerable interest in increasing the bang we get for the public education buck or, put differently, in increasing the productivity of public schools and educational institutions in the United States.

Determining the level of funding needed to accomplish educational goals has become a central focus of the policy debate in education. Some contend that schools can “do better” by using resources more effectively say, by moving school personnel out of administrative positions and into classrooms, or by focusing curricula on academic fundamentals such as reading and mathematics rather than arts or physical education. Others argue that better performance requires additional resources to decrease class size, attract more qualified teachers, or provide teacher professional development. Public pressures have spawned a plethora of research examining the relationship between educational resources and student performance, shedding light on the production of education and, at the same time, highlighting substantial gaps in our knowledge. Can we raise test scores by reducing class size? Would it be better (or cheaper) to increase the number of teacher aides instead? Should we invest in professional development for teachers or lengthen the school day (or year)? These are, essentially, questions about the productivity of school resources.

Borrowing from the economics of manufacturing, research on productivity in education often focuses on estimating an “education production function” that links school inputs to educational outcomes and identifies the impact of changes in inputs (e.g., teachers) on student outcomes (e.g., achievement as measured by test scores). While the application of economic tools has been fruitful, the productivity of schools is considerably more complicated than the relatively simple analogy to manufacturing suggests. In manufacturing, "productivity" is typically defined and measured as the amount of output produced per unit of input. This seemingly straightforward concept is riddled with complexities when applied to education. For instance, questions surround the appropriate outputs of education. What should schools produce and how should we measure the quality and quantity of the “product”? Should we focus on basic functional competencies or critical thinking skills? Will standardized tests suffice to measure progress and if so, what sort of tests? What about civic responsibility, economic self-sufficiency, cultural awareness, and social and economic mobility? How do we account for those outcomes? Answering these questions extends well beyond the domain of research and requires a broader social consensus on the fundamental goals of public education. The absence of such a consensus complicates the measurement of productivity in education.

Identifying and measuring inputs is equally challenging. What are the inputs to the production of education? The spectrum of educational resources or inputs is wide and varied across communities. While purchased inputs, such as teachers, school buildings, or books are easily identified, nonpurchased inputs, including donated goods and services, are less obvious but no less critical to student success.

Of course, measuring inputs and outputs is only the first step in understanding the factors that drive differences in productivity across schools and school districts and, ultimately, identifying ways to make improvements. Put differently, researchers aim to peek inside the “black box” of education and examine the technology of production to understand what schools do with their resources, how these choices influence the relationships between inputs and outputs and, ultimately, what changes can be made to improve that relationship. Thus, research on productivity in education might consider the implications of different choices related to curriculum (e.g., whole language vs. phonics), the organization of schools (e.g., K–8 or K–5 and 6–8),
the size of schools and classes, policies toward teachers (e.g., hiring, tenure, and compensation policies), and so on. As might be expected, operationalizing these constructs and developing appropriate measures to study them present theoretical and empirical challenges.

Finally, researchers have struggled with understanding the role that students themselves play in the production process. Clearly, student differences in ability, level of motivation, and aspirations are important determinants of educational outcomes. These student characteristics may be shaped by school policies or programs at the same time that they influence the kinds of policies or programs that the schools adopt. Thus, disentangling the causal pathways that link resources to outcomes is complicated by the responses of students (and parents, teachers, and others) in ways that may complement or hinder the intentions of school policies.

The goal of this chapter is to describe and discuss the theoretical foundations and empirical contributions of research on productivity in education. We aim to provide a sense of the depth and breadth, the complexity and challenges of this field of research. At the outset, we recognize that "productivity in education" means different things to different people, but is fundamentally concerned with the quantity and quality of educational outcomes that result from a given investment of resources. The conceptual framework underlying much of the research linking educational resources and student outcomes is most fully detailed in the education production function. Developed by economists drawing on models of production in agriculture and manufacturing, the concepts formalized in the education production function have been applied in productivity studies across a broad range of disciplines and research traditions. In the section that follows, we begin our discussion with the economist’s notion of the education production function, and discuss the key empirical debates and conceptual issues. We then provide a brief description of the various research traditions that have examined the relationship between educational resources and student outcomes. We conclude with a discussion of the role of research on productivity in education in supporting multiple, competing social goals of public education in the United States.

**The Education Production Function**

Education production functions have been used to address a set of fairly straightforward policy questions:

- How much will outcomes improve if we increase resources by some amount (e.g., how much will test scores rise if we decrease class size by 10 percent)?
- Can schools deliver better outcomes with existing resources or will better outcomes require more resources (e.g., can we increase performance by eliminating waste or improving efficiency)?
- How much more will it cost to improve outcomes by a desired amount or to a desired level (e.g., what will it cost to meet standards on state proficiency tests)?

Answering these questions requires an understanding of the production process—the inputs and outputs and the relationship between them—and also an understanding of the prices that schools must pay for these inputs (e.g., teacher salaries) and how efficiently schools (or school personnel) use their resources. We begin with a discussion of outputs in education and turn quickly to our central concerns—defining and measuring school inputs; inefficiency and waste; and a more formal mathematical presentation of the production function. We conclude with discussion of productivity and social values.

**School Outputs**

The starting point for any study of productivity in education is the definition and measurement of outcomes. In general, the purpose of education is to produce individuals who can contribute to the economic, political, civic, social, and cultural institutions in our society. As such, we expect high school graduates to have acquired a wide range of competencies, skills, and personal qualities. This goal is consistent with that of early proponents of public education, including Horace Mann and Thomas Jefferson (Jacobsen & Rothstein, this volume). Further, it reflects international efforts to identify the array of "key competencies that contribute to a successful life and a well-functioning society" (Rychen & Salganik, 2003). Research on the productivity of
education resources, however, typically focuses on a narrow set of measures driven, in large part, by the quantity, accessibility, and quality of available data.

Some studies estimate the effect of education investments on long-term outcomes such as employment and earnings (e.g., Betts, 1996; Bishop, 1991; Card & Krueger, 1996, Chetty et al., 2011), civic participation (Dee, 2004), and noncognitive outcomes (Dee & West, 2011). Using earnings as the outcome has a particular appeal because it delivers an estimate of the impact of education resources in dollars, making it easy to estimate the rate of the return on the investment, which then can be compared with the rates of return associated with other sorts of public investments (e.g., health care, social services). The difficulty with using earnings, or any distal outcome measure, however, lies in the challenges of isolating the specific effect of the public school from other factors, including labor market conditions and the availability of college opportunities, that may shape these outcomes in the time between graduation from public school and the measurement of the outcome.

The more typical alternative is to focus on shorter-term outputs such as student achievement measures (generally in math and reading for elementary schools and math and science for high school) and high school graduation rates. While this set of commonly used output measures reflects a narrow view of the broad range of purposes for public education, these measures are consistent with the current emphasis of accountability systems on student test scores (see also Jacobsen & Rothstein, this volume). And, although there continues to be some policy interest in understanding the determinants of test score levels, measuring educational output requires controlling, in some way, for the student’s prior educational performance or attainment. One way to estimate the value added by a particular resource to student performance is to measure output as the gain in student performance (that is, the increase in test scores between years). An alternative is to include prior performance as a control variable in regression analysis, which is particularly attractive for studies that focus on nontest score outcomes such as graduation rates.

**School Inputs**

As described earlier, education output reflects not only the impacts of school inputs but also the influences of families, communities, and individuals. Consequently, isolating the specific contributions of school inputs such as class size, teachers, or computers—the focus of much research on productivity in education—requires accounting for the contributions of peers, family, and communities. These outside-of-school influences are interesting and important in their own right, and a significant body of research examines the impact of home- and community-based inputs including health care, parenting practices, and home environment on outcomes. (See Rothstein, 2006, for a discussion of the role of these factors as determinants of student performance.) That said, most of the policy research—and the policy debate—focuses on school inputs, since the public sector bears much of the responsibility for ensuring equity and efficiency in provision, and it wields considerable power in determining those inputs. Thus, while family practices undoubtedly shape student performance, school districts have little ability to influence out-of-school practices and have much wider latitude in changing school inputs.

School inputs into the education process include a wide range of resources, both purchased and donated—teachers, guidance counselors, school buildings, buses, books, computers, paper, and so on—and the quality and character of these inputs varies widely. Thus, measuring the quantity and quality of inputs can be difficult, and, as a result, data of this kind are relatively scarce. Consequently, many researchers have relied upon spending data, which can be viewed as providing a summary measure of the amount of inputs used, and which easily yield estimates of the improvements associated with each dollar spent.\(^1\)

For a variety of reasons, however, measures of spending fall short of representing the quantity and quality of resources available to students. As a result, studies examining the effect of per-pupil expenditures on student performance and eventual labor market outcomes, taken together, have been frustratingly inconsistent in their findings. While Hanushek (1986) finds little consistent evidence that school expenditures have a statistically (much less substantively) significant impact on test scores, Hedges, Laine, and Greenwald (1994) are more sanguine in their reading of the evidence, finding that spending may, in contrast, be important. By calling into questions Hanushek’s “vote-counting” methodology of tabulating research findings, Krueger (2002) demonstrates that other approaches to weighting studies lead to a more consistent and positive story about the effect of resources on student achievement.\(^2\) Despite the intense policy debates and considerable heat generated about whether or how much “money matters,” few people would argue that money never matters. For
instance, it is hard to imagine that anyone would argue seriously that cutting spending back by 50 or 75 percent would have no impact on educational outcomes. On the other hand, it is equally hard to imagine a serious argument that money is always the answer or that doubling spending would necessarily lead to a twofold improvement in student performance.

Ultimately, the failure to reach consensus may reflect inadequacy in the framing of the question. As described above, the relationship between dollars spent and output realized is not direct. Instead, money “matters” because (or to the extent that) it changes the quantity and/or quality of inputs used in the production of education. Thus, recent research has moved away from examining the level of spending to focus on understanding the impact of what the money buys (e.g., school buildings, school leadership, instructional resources, computers, teachers) and on understanding why (or how) prices or costs vary. Much of this work focuses on three key determinants of school spending and outcomes: class size, teacher qualifications, and teacher salaries.

While debate continues, new evidence on the impact of class size and teacher quality on student performance is mounting and consensus may be emerging. New methods yield increasingly convincing evidence that class size matters, improving upon existing studies which had largely relied upon the variations in class sizes across schools in U.S. districts to estimate the impacts. Angrist and Lavy (1999) investigated this issue using data on Israeli public schools where significant differences in class sizes emerge because of small, essentially random, differences in the number of students in a grade up to the maximum class size of 40 students. These researchers found significant and substantial increases in test scores of reducing class size. Equally important has been the evidence from the Tennessee STAR class size experiment. Researchers randomly assigned students and teachers in grades K–3 to small classes (13–17 students), regular classes (22–26 students), or regular classes with a full-time teacher’s aide in addition to the classroom teacher to study the effect of the assignment on student test scores. This sort of randomized experiment has been heralded by some as the “gold standard” of social science methods.

Despite the hope that the STAR experiment would settle the class size debate, controversy has surrounded the interpretation of the STAR results. Many interpret the results as indicating a positive effect of reducing class size, with the largest effects for minority students (see Finn & Achilles, 1999; Krueger, 1999), and some have demonstrated that the effects persist over time (Nye, Hedges, & Konstantopolous, 1999). Hanushek (1999) is unpersuaded, arguing that the implementation of the experiment was flawed—leading to unmeasured differences between the students in small and large classes, for example—and that the impact of reduced class size in the STAR experiment is limited to large reductions in kindergarten and possibly first-grade classes. Further, STAR provides evidence on the impact of reducing class size when students are assigned to classes randomly rather than purposefully to match learning styles or abilities, create synergies, or avoid conflicts as is practiced by many schools. That said, Hanushek (2002) acknowledges that “class size reductions are beneficial in specific circumstances—for specific groups of students, subject matters, and teachers” (p. 38). At the same time, there is a growing recognition that class size reduction is not a magic bullet for improving academic performance. As demonstrated by California’s large-scale class size reduction initiative, reducing class size requires an increase in the need for teachers (and classroom space), which may carry unintended consequences for schools, even while test scores may rise overall (Jepsen & Rivkin, 2002). Thus, while the class size debate continues, it seems to have become more nuanced, focusing on when, where, and for whom class size reduction makes most sense. (See chapters by Krueger, Hanushek, and Rice in Mishel & Rothstein, 2002, for more information.)

While teachers are clearly the most important educational resource provided to students and research increasingly finds that the quality of teachers matters significantly to student performance (Rivkin, Hanushek, & Kain, 1998; Sanders, 1998), evidence on the relationship between specific teacher qualifications and student performance is mixed and inconclusive, undermining policies that rely on teacher qualifications as indicators of teacher quality (Rice, 2003; Goldhaber, this volume). Alternatives to improve teacher quality include policies focused on relaxing requirements for entry into the profession (“teacher gateway policies”) and policies aimed at improving the effectiveness of existing teachers (“teacher workforce policies”) (Goldhaber, this volume).

Much of the existing research on teacher effectiveness and qualifications focuses quite narrowly on the contributions of an individual teacher to the gains made by the students in his or her own current classroom. In contrast, other school personnel may also play a role in determining the performance of single student.
their study examining the effect of teacher education and preparation on student performance in high school math and science, Monk and King (1994) distinguished between different levels of teaching resources in a school: (1) the “proximate” teacher who is the currently assigned instructor in a particular subject; (2) the immediate past teacher in the subject; (3) the set of previous teachers a student had had in a particular subject over some period of time; and (4) the set of all teachers in a school who teach courses in a given subject. One argument for pursuing the school-level effects of teacher qualifications is that teachers learn from one another, so any negative effect associated with having a low-quality teacher might be reduced if there are other teachers who are supportive, more knowledgeable, and more skilled. A second theory is that students interact with and learn from teachers other than those to whom they are directly assigned. Using national data, the researchers found that it is the cumulative effect of the set of teachers that a student has had over time that affects student achievement, particularly in math. This study demonstrates the importance of moving beyond the narrow linking of each student to one teacher in productivity analyses to consider the role of other personnel—say subject area teachers, teaching support personnel, guidance counselors, or principals—in determining student performance.

In addition, recent research examines differences in the costs of purchased inputs, particularly teachers, and focuses on what a dollar buys. This research has been motivated, in part, by an interest in explaining how and why a dollar spent in one location has a different impact on outcomes than a dollar spent elsewhere, or how and why a dollar spent in one year may have a different impact than a dollar spent some years later. Ultimately, the key question is whether these differences can be explained by differences in the price of acquiring inputs—due to differences in labor markets, for example—or by differences in the productivity of these inputs.

The cost of educational inputs varies geographically due to differences in local labor markets, house prices, transportation costs, energy prices, and the like. For example, the cost of building a new school is higher in San Francisco, where property values and construction costs are relatively high, than in Cleveland, where these costs are more modest. Further, the high cost of housing and competitive labor markets make it likely that a higher salary will be needed to hire a well-trained teacher in New York City than in a more affordable place like Tallahassee. Despite the intuitive appeal, measuring cost differentials is difficult due, essentially, to the scarcity of geographically detailed data on the prices and quality of inputs purchased. Nonetheless, some headway has been made and various estimates of geographic cost differentials in education, mostly focused on teachers, now exist. A prime example is the Geographic Cost of Education Index (GCEI) developed by Chambers (1998), which adjusts salary differentials for differences in worker and workplace characteristics (see Duncombe, Nguyen-Hoang, & Yinger, this volume). States have also implemented their own cost-of-education indices to compensate school districts for geographic variations in cost—mostly associated with teacher salaries—that result from factors beyond the control of school districts (e.g., Taylor et al., 2002). (For a relatively recent example, see Allegretto, Corcoran, & Mishel, 2004.)

As noted above, related research has addressed changes in costs over time. There may be some consensus among economists, if not more widely, that the education sector may suffer from what is called Baumol’s cost disease. According to this view, labor-intensive sectors, such as education, will experience lower productivity growth relative to other sectors such as manufacturing that rely more on capital and machines than on human interactions for production. The implication is that technological change in making cars and computers will make them cheaper over time (controlling for quality), while technological change in personnel-intensive industries including education, nursing, and hospital care will be slower, making the products of these sectors relatively more expensive over time.

Finally, it is important to consider the role of students in their own education, and the way differences in their motivation and behavioral responses shape the output of schools. Put simply, students (and parents) are not passive translators of education inputs into outputs. Instead, the process by which the inputs are translated into outputs depends critically upon behavior, and ignoring this response may confound our understanding of productivity in education. If, for example, schools provide more books for students to read at home and parents or students respond by decreasing the number of books acquired from the library, then the impact of the school book program may be small or insignificant even if having more books at home increases student reading ceteris paribus. (See Todd & Wolpin, 2003, for more on this point.) Unfortunately, much of the research on productivity yields average effects that typically do not account for the differences across students in
motivation and behavioral responses to resources.

A Note on Inefficiency and Waste

We should note that in our discussion thus far we have devoted little attention to the possibility that resources are wasted or used inefficiently in public education. Considerable research recognizes this possibility (e.g., Hoxby, 1995; Duncombe & Yinger, 1999). One reason that schools may be wasteful or inefficient is the possibility of a mismatch between the goals of the school personnel and the goals implicit in the formulation of the production function. For instance, the school may be aiming to maximize graduation rates while policymakers and analysts are gauging efficiency in terms of producing SAT scores as an output. If these two outputs are highly correlated, then the differences may matter little. If not, this mismatch between goals and measured outputs can matter a good deal. A second reason schools may be inefficient is that they may be constrained by policies or institutions, such as work rules for faculty and staff or building construction and maintenance, that prevent them from making the best use of their resources in ways that are unobserved by the researcher. A third possibility is that inefficiency arises because of poor motivation, malfeasance, or inadequate information about the levels of outputs. Or, special interest groups might be able to influence school policies in ways that benefit themselves at the expense of others. As an example, Rose and Sonstelie (2007) argue that the premium paid to experienced teachers is inefficiently high because of the influence of the teachers union. Clearly, the prospect of improving performance by eliminating waste or reducing inefficiency is an attractive policy option, suggesting that we can get more without additional resources. Unfortunately, the evidence on this remains controversial, although research on measuring efficiency and on policies that reduce waste and enhance efficiency is promising (see, e.g., Stiefel et al., 2005)

The Education Production Function: A Mathematical Treatment

An education production function typically takes the following general form:

\[ A_{it} = g(F_{it}, S_{it}, P_{it}, O_{it}, I_{it}) \]  

where

- \( A_{it} \) is a vector of educational outcomes for the \( i \)th student at time \( t \)
- \( F_{it} \) is a vector of family background characteristics relevant to the \( i \)th student at time \( t \)
- \( S_{it} \) is a vector of school inputs relevant to the \( i \)th student at time \( t \)
- \( P_{it} \) is a vector of peer or fellow-student characteristics relevant to the \( i \)th student at time \( t \)
- \( O_{it} \) is a vector of other external influences (e.g., community) relevant to the \( i \)th student at time \( t \)
- \( I_{it} \) is a vector of characteristics of the \( i \)th student relevant at time \( t \)
- \( g \) is the transformation that captures the technology linking them

This formulation is conceptually comprehensive in that it includes a wide array of possible inputs (school, community, family, peer, and individual across time) and outcomes (achievement as well as other more difficult to measure outcomes such as civic competencies and cultural tolerance). To be clear, these vectors may include variables that measure characteristics from prior periods, including, but not limited to, the student’s prior education experiences, or home environment in early childhood. Such variables are designed to account the cumulative nature of the education process. A common way for researchers to control for these prior experiences is to include only the student’s previous academic performance on the ground that this measure sufficiently captures all their effects. This assumption is not fully satisfying, and a growing body of research treats prior year experiences with greater nuance (e.g., Todd & Wolpin, 2003).

When prior year achievement is included, equation 1 can be interpreted as a value-added model. In this case, the estimated coefficients indicate how each of the explanatory variables in the current period influence achievement, controlling for the achievement that the student brings to the classroom. An alternative way of
specifying a value added model is to substitute the gains in output (e.g., the change in test scores from one year to the next) for the level of output in equation 1. While these models are flexible in principle, limitations in methods and the availability of data on both inputs and outputs can be significant in practice.

In many applications, the production function is estimated using regression analysis and assumes a linear relationship between the inputs and the outputs:\(^2\)

\[
A_{it} = b_0 + b_1 F_{it} + b_2 S_{it} + b_3 P_{it} + b_4 O_{it} + b_5 I_{it}
\]  

Thus, the regression yields estimates of \(b_0\), an intercept, and \(b_1, b_2, b_3, b_4,\) and \(b_5\) which measure the increase in the outcome \(A_{it}\) that results from a one unit change in the corresponding input \(F_{it}, S_{it}, P_{it}, O_{it},\) and \(I_{it}\), respectively.\(^5\) Notice that the estimates of \(b_1, b_2, b_3, b_4,\) and \(b_5\) are key productivity measures, capturing what economists call the marginal product of the inputs.\(^5\) The parameter \(b_2\) is of particular interest for education policy-makers because it measures the change in output that is due to a change in school inputs. As an example, if we specify \(S_{it}\) as spending per pupil, then \(b_2\) will provide an estimate of the impact on outcomes of increasing spending per pupil by a dollar.

Estimating a production function requires identifying a set of schools that are sufficiently similar that we can summarize their production process using a single production function. The schools can differ in the level or mix of the inputs they use, their output level, community factors, and so on, but in estimation we assume that the parameters of the production function that summarize the average relationship between variables is the same for all. Often production functions are estimated using data on all of the schools in a particular district or state, or using data on a sample of schools from across the nation. The sample is typically limited further, distinguishing between elementary, middle, and high schools, and separating schools serving primarily special education students, or schools in their first year of operation. Further analyses often focus on particular populations of students (e.g., low income, special education) or subjects (e.g., mathematics, reading) within those schools.

Using several years of data for the same set of schools is particularly attractive because it allows the specification to include a school fixed effect (a school-specific intercept) which means the impact is estimated using the differences in inputs and outputs experienced by the same school over time and controls for any unobserved, time-invariant features of schools that might not be fully captured by the other variables. Although this method controls for persistent differences among schools, it does not control for changes over time in factors such as school leadership or school curricula. That said, longitudinal analyses can be very powerful and can provide insight that would be difficult to disentangle using only cross-sectional data.

This framework clarifies that increasing output can be accomplished in two very distinct ways: (1) by changing the quantities of the inputs—increasing the quantity of at least one input, potentially substituting for another input, and (2) by increasing the marginal products of the inputs.\(^2\) This distinction is important and often obscured in policy discussions. As an example, the first method might involve increasing the number of teachers while decreasing the number of paraprofessionals. An example of the second approach would be to introduce new instructional practices that help students do better given existing inputs. Research on productivity in education explores both kinds of interventions.

Productivity researchers have begun to recognize that numerous levels of decision making simultaneously occur at any point in the educational process, and all have potential impacts on the productivity of the entire enterprise. In other words, the complexity of schooling is manifest in the interlocking and nested nature of its production function. The educational process is a function of student and family characteristics as well as policies and practices at multiple organizational levels: students are nested in classrooms, classrooms are nested in schools, and schools are nested in districts. Further, the education system is nested in a broader social and economic policy context. These influences are all operating simultaneously to create the whole educational experience for each pupil. The current challenge for productivity research is to incorporate the various actors in and levels of the process, and to disentangle their effects on pupil outcomes.

While the education production function as promoted by economists can serve as a powerful conceptual and analytic tool for understanding the relationship between inputs and outcomes, this approach has not been fully embraced by the education community. Critics have raised concerns about its emphasis on average effects across schools, its lack of attention to the process of educating students (i.e., what happens inside the "black
box”), and its apparent neglect of the role of students in their own education. New research methodologies and data collection that allow education production function studies to address these concerns and/or pair production function studies with qualitative analyses that provide depth and context are attractive and likely to be given greater attention in the future.

Characterizing the Research on Productivity in Education

While research on productivity is often considered to be part of the discipline of economics and much of the research on productivity in education per se has been conducted by economists, scholars from multiple research traditions have explored the relationship between inputs and outcomes in education. While many draw on the conceptual framework of the education production function—implicitly if not explicitly—a significant body of education research draws on qualitative methods to complement the quantitative. These studies employ designs aimed at addressing the inherent trade-offs between the generalizability offered by large-scale quantitative studies and the in-depth context-specific findings of smaller scale qualitative research. From a methodological perspective, the most promising approach may involve mixed methods, where complementary approaches compensate for the weaknesses inherent in each (Rice, 2001). Ladd and Hansen (1999, pp. 139–140) offer three “lenses” to characterize this “large and unwieldy body of research” (Monk, 1990, p. 315).

The first lens, input-output research, uses the production function framework described above to isolate a causal relationship between school resources and student outcomes, usually measured by test scores. This line of research dates back to the 1966 Equality of Educational Opportunity Study, which concluded that schools have little influence on student achievement, independent of family background (Coleman et al., 1966). With the availability of better data systems and the development of more sophisticated analytic methods, the past four decades have seen these studies evolve into increasingly sophisticated statistical analyses. While this research has examined the impact of multiple resources (e.g., school facilities, computers, and technology), much of the research in this area, as described above, has focused on class size, teacher qualifications, and teacher salary.

A second lens for characterizing research on education productivity is the institutional perspective. This body of work includes studies that examine how broad education policies can influence the will and capacity of schools to use their resources more productively. Much policy interest has been paid to the possibility of improving outcomes by intensifying the “pressure” placed on schools and school districts through incentives embedded in standards-based accountability and market-based reforms. High-stakes accountability policies that rely heavily on student testing to promote better performance within traditional public education systems have become pervasive across education systems throughout the nation. These systems often link the results of standardized tests to rewards (e.g., pay incentives) or sanctions (e.g., school reconstitution) for schools and school personnel. The underlying logic is that such systems improve outcomes by inducing schools to eliminate waste which may arise because of malfeasance, lethargy, or ineptitude (of myriad kinds); because the absence of clear standards has led to a misalignment between the goals of the schools and those of the state; or because innovative educators have not been motivated to find new and better ways to accomplish those goals. In 2002, the No Child Left Behind Act made such a system national, but many states had enacted some sort of accountability system prior to that year, providing researchers the opportunity to gain insight into the effectiveness of these reforms. While theory suggests that strengthened incentives in high-stakes accountability systems lead to improved performance (Hanushek, 1994; Koretz, 1996; Malen & Rice, 2004), the effectiveness of these policies depends on a number of design elements including the measurement of school performance, the treatment of highly mobile and disabled students, the assessment of student subgroups within schools, the accountability time period, the capacity of schools, and the breadth and focus of the accountability system (Figlio & Ladd, this volume; Malen et al., in press; Rice et al., in press). Although research on performance incentives in the United States is limited, the growing evidence has provided little support for the proposition that pay-for-performance systems significantly improve performance (Springer, 2009; Fryer, 2013).

Another institutional approach aimed at improving productivity is to introduce market-based reforms that expand the choices of students and families beyond those available in traditional public education systems.
Market-based reforms increase competition for students between schools and create incentives for schools to increase performance to attract more and/or better students and the funds they bring to the school. Theoretically, market-based reforms also free schools from the bureaucratic demands of a public system, allowing them to be more innovative and productive. These options—including charter schools, public funding for private schools, and private management of public schools—are sometimes components of high-stakes accountability policies, so the two approaches are not mutually exclusive. Theories about the power of markets to improve school performance date to the 1950s, in the arguments advanced by Milton Friedman. Despite the promising theoretical basis for market-based reforms as a mechanism for improving productivity (Friedman, 2002), research indicates that the effectiveness of these strategies is contingent on a range of design factors. Perhaps most important are the effects of these policies on conventional public schools and on the broad set of purposes for public education. Existing research suggests cautious optimism (Gill & Booker, this volume). Both standards-based accountability and market-based reforms can be understood as efforts to induce schools to produce the maximum possible output given their resources and existing education "technology," or to encourage school systems to develop new technologies that allow them to produce more with existing resources. (In the parlance of economics, these might be viewed as aiming to move schools onto the production possibilities frontier or to shift the frontier.)

A third lens for viewing the relationship between education resources and outcomes are studies of educational practices. Studies of educational practices look inside the "black box" (Ladd & Hansen, 1999; Rice, 2001) to examine the inner workings of schools and school systems with the goal of identifying the organizational conditions, educational resources, and instructional strategies associated with high performance. Anthropologists, sociologists, psychologists, and others have used a range of research traditions to empirically identify factors associated with effective schools that might be replicated in other environments to improve performance. These studies—often characterized as school effects studies, process-product studies, and outlier studies—employ qualitative and quantitative methods to identify key factors associated with higher-than-expected performance, given the demographic characteristics of the school. Often focusing on urban contexts, researchers select schools serving similar student populations, but differing significantly in student achievement, with the goal of identifying characteristics that distinguish effective from ineffective schools (e.g., Brookover et al., 1979; Edmonds, 1981). Taken together, this body of research has generated a five-factor model of effective schools: strong building leadership, shared instructional focus, safe and orderly school climate, high expectations for student learning, and data-based decision making (Edmonds, 1979, 1981). While this model was influential in education policy and reform efforts in the 1980s (Teddle & Stringfield, 1993; General Accounting Office, 1989), scholars have identified a number of limitations associated with this body of work including concerns about the validity of outcome measures, instability of effects over time, lack of attention to contextual factors, and a heavy reliance on student and teacher perceptions (Good & Brophy, 1986; Purkey & Smith, 1983). Further, since the analyses are based upon a sample of schools chosen because of their higher-than-expected outcomes, it is possible that selection bias has, at least to some extent, tainted the results.

The next generation of studies of educational practices focuses on approaches to school reform that are consistent with key principles emerging from this body of research, such as the importance of high standards for all students, the need for good curriculum and teaching, and the role of school staff in school improvement efforts (Smith et al., 1996). Following years of top-down reform efforts, education policies based on effective schools research emphasized schools as the unit of change (Tyack & Cuban, 1995). For instance, school-based management (Malen, Ogawa, & Kranz, 1990; Tyack, 1993; Wohlstetter & Odden, 1992) that prioritizes school-level decision making, and whole-school reform models (Wong & Wang, 2002) that emphasize the school as the unit for comprehensive reform became popular strategies. However, studies examining the effectiveness of these approaches were undermined by uneven implementation, insufficient data and methods of analysis, and the lack of attention to contextual factors (Summers & Johnson, 1996; Bodilly, 1998; Wong & Wang, 2002). Further, the effects of these reforms need to be considered in light of their costs to make determinations about productivity (Rice, 2002).

More recent studies employ research designs aimed specifically at identifying the causal effect of school reforms on outcomes and probing the “black box” with complementary analyses. For instance, research into the efficacy of small high schools has used instrumental variables (Schwartz et al., 2013; Barrow, Claessens, & Schanzenbach, 2013) and the randomized assignment of students by school districts allocating scarce seats at
Productivity in the Broader Context of Social Values

Underlying this chapter is the assumption that productivity research can help guide policy-makers in their efforts to improve the allocation and use of educational resources. Specifically, productivity research provides information about the relationships between educational inputs and outcomes, and when this type of information is coupled with data on costs, it can guide policymakers seeking to identify the most efficient policy alternative. Clearly, this approach has much to offer in the current policy environment that gives high priority to accountability and efficiency, particularly if the outcomes measured are consistent with the goals sought. However, efficiency is not the only goal underlying the financing of public education. The pursuit of equity has also been an enduring concern of education finance policy and research.

While the link between productivity research and the goal of efficiency has always been apparent, the connection between productivity research and the goal of equity has grown more direct as the concept of equity has evolved from a focus on educational inputs (e.g., equal dollars per pupil) to an emphasis on outcomes. This shift is most apparent in the adequacy movement, which promotes funding systems that provide "sufficient resources to ensure students an effective opportunity to acquire appropriately specified levels of knowledge and skills" (Guthrie & Rothstein, 2001, p. 103; Koski & Hahnel, this volume). Adequacy-based reform requires clearly specified and measurable academic proficiencies expected of all students, and information on the educational inputs required to provide all students the opportunity to meet those standards. So, specifying and evaluating adequacy requires knowledge of the education production function, that is, the process by which inputs are systematically transformed into desired outcomes.

In addition, funding an adequate system of education requires information on the cost of providing educational opportunities for all students to meet standards. This focus on the costs of an adequate education introduces questions about the resources required to educate students with a variety of learning styles and needs across different kinds of communities. The work on this topic reveals the importance of efficiency as a prerequisite for adequacy determinations (Duncombe & Yinger, 1999). An adequate funding system should offer additional aid to districts that face higher-than-average education costs due to factors outside of their control, but not to districts whose costs are high due to inefficiency. Again, knowledge of the production process is needed to make these determinations.

While progress is being made in the study of productivity in education and, specifically, using education production functions, much remains to be learned about what resources, in what quantities, and under what circumstances lead to learning. Perhaps the greatest promise of adequacy-based school reform is that it will promote a greater accumulation of knowledge about the relationship between school inputs and student outcomes. This task may involve efforts that both improve the quality of production function research and expand the range of research brought to bear on these matters. The research on class size reduction, for example, suggests that we need to draw on a range of research designs, including but not limited to production function studies, to gain a clear understanding of these relationships and whether and how context shapes them. The key to improving student achievement through specific investments lies in reaching a more refined understanding of the resources themselves, for whom they matter most, under what circumstances, and at what cost. Such work is critical to the promotion of efficiency, equity, and adequacy in public education.

Notes

1. Public expenditures on education are substantial. In 2003, more than $511 billion, or 4.7 percent of the U.S. gross domestic product, was dedicated to elementary and secondary education. The average total expenditure per pupil ranged from $5,969 in Utah to $14,419 in Washington, D.C., with a national average of $9,299 (U.S. Department of Education, 2005).

2. Krueger’s (2002) critique of Hanushek’s methodology centered on the way in which the various studies were weighted in Hanushek’s analysis. Essentially, Hanushek labeled each estimate of an effect as a "study," so that one article could have several estimates, or studies, that
are factored into Hanushek’s count of positive, negative, or statistically insignificant (positive and negative) effects. Krueger argues that Hanushek’s approach weights the various studies by the number of different estimates of the effect of a particular variable they include. Further, he contends that studies that report negative or statistically insignificant findings are more likely to include more estimates than those that find statistically significant positive effects.

3. This specification follows Levin (1976). A similar specification can be found in Hanushek (1986).

4. Technically, an econometric specification of (2) would include an error term recognizing the potential for random events to influence the educational outcome.

5. In principle, this interpretation is correct only if the model is fully and correctly specified such that the estimates are unbiased.

6. In principle, all inputs should have a positive marginal product. That is, increasing the level of any one input will raise the output produced. So, increasing the number of teachers or books, say, always leads to an increase in output, even if it’s small. This is less restrictive than it may seem and essentially reflects an underlying assumption that there is, in the language of economics, “free disposal,” meaning that any input that does not make a positive contribution can be disposed of at no cost.

7. The first implies movement along the production function; the second implies changing the production function.

8. In some sense, these kinds of reforms aim at getting schools back on the production frontier, rather than moving along the frontier.

References


Duncombe, W. D., & Yinger J. (1999). Performance standards and educational cost indices: You can’t have one without the other. In H. F. Ladd, R. Chalk, & J. S. Hansen (Eds.), Equity and adequacy in education finance:


Cost-Benefit Analysis and Cost-Effectiveness Analysis

CLIVE R. BELFIELD

Introduction

Cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA) are methods for evaluating interventions, reforms, and policies. CBA and CEA supplement commonly used evaluation methods related to effectiveness and efficacy by adding an economic component to the evaluation. There is a strong case for some form of economic evaluation across most government programs, including those in the education sector. Every education program requires resources, and the impact of the program may depend on the quantity of resources applied. For example, a mentoring program for a school of 100 students may have five counselors or 10 counselors; an after-school program may operate in a school playground or have extensive, dedicated facilities; and a teacher professional development program may run for one week or six weeks. In each example, the more resource-intensive program might be expected to improve education outcomes to a greater extent. CBA and CEA incorporate resource use directly into an evaluation to test for efficiency—i.e., whether the resources are being used in the most effective way to generate the greatest possible improvement.

There are two senses of efficiency: one refers to investments that yield outcomes that society values (allocative efficiency); the other refers to using inputs in the most effective way to produce the desired outcomes (technical efficiency). CBA indicates whether an intervention is (allocatively) efficient by looking at whether the dollar costs of the resources needed for implementation are less than the dollar benefits that the intervention produces. CEA indicates which intervention is the most (technically) efficient by looking at which intervention generates the biggest improvement in educational objectives per dollar spent.

CBA and CEA apply the basic economic concept of opportunity cost to educational interventions. All interventions require resources, and the appropriate way to evaluate them is by comparing them to the next best use of equivalent resources. For example, if a school is choosing across a range of curricular interventions to improve outcomes for struggling readers, it should look at the resources required to implement each intervention in conjunction with its effect on reading. Then, the school should choose the most cost-effective intervention—i.e., the one that improves reading outcomes by a given amount at the lowest cost (or by the largest amount for a given cost). If a state is considering expansion of its college programs, it should compare the costs of the programs to the returns from graduates' higher earnings, as well as other benefits. By applying the concept of opportunity cost, these analyses take account of the fact that resources are scarce and that programs should be chosen based either on how effectively they use the available resources or on what return is made on the investment.

CBA and CEA are companion forms of analysis: they share many common principles and require similar data. Both include analysis of how resources are used to implement education programs. But the two methods differ significantly in how they analyze the outcomes of a given program. CEA relates costs to effectiveness, where the latter is denominated using an education metric (such as a graduation rate or reading gain). The results of a CEA are presented as a cost-effectiveness ratio where the total costs are divided by the resulting outcomes. The intervention with the lowest ratio is the most cost-effective, that is, it is "buying" outcomes at the lowest price. Thus, CEA helps identify which intervention is the most efficient at achieving a particular set of desired effects (relative to other interventions). CBA requires an additional step, which is to translate measures of effectiveness into monetary values. After this translation, a benefit-cost (or cost-benefit) ratio can be calculated with both the numerator and denominator expressed in monetary units. The program with the highest benefit-cost ratio is the most efficient, that is, it yields the most dollars back per dollar spent on the
program. The advantage of cost-benefit analysis is that it determines whether an intervention is worth implementing at all, that is, whether the benefits exceed the costs (or the benefit-cost ratio exceeds one). The disadvantage is that CBA requires considerably more data and analysis.

CBA and CEA are intended to help education policymakers and professionals make decisions. As educational budgets are constrained and not all programs can be funded, it is critical for decision makers to have information on which interventions are the most cost effective. CBA and CEA can be readily applied in education research to help decision makers evaluate many interventions, reforms, and policies. Their application is especially useful in the public sector where there is no obvious driver (such as the profit motive) to ensure efficiency. Given their differences, an evaluator would typically apply either CBA or CEA and not both together. If a decision maker has already specified the educational objective and is seeking the best way to achieve this objective, cost-effectiveness analysis is sufficient. If the decision maker is uncertain that a program is worth the investment, then cost-benefit analysis is needed.

However, cost-benefit and cost-effectiveness ratios do not compel decision makers to choose a particular policy. Other factors, such as political considerations or budget constraints, may be influential. For example, interventions where the costs exceed the benefits or those with a high cost-effectiveness ratio may still be chosen if they serve important social goals (such as giving all students an equal opportunity to learn or helping disadvantaged students). The presumption is that, absent any policy constraints, alternative policy goals, or equity considerations, the intervention with the lowest cost-benefit or cost-effectiveness ratio should be preferred. But it is still up to the policymaker to make this determination.

The method of cost-effectiveness analysis, at least in its application for education research, was developed by Henry Levin in his seminal textbook of 1983, with a second edition published in 2001 with coauthor Patrick McEwan. (The general method of cost-benefit analysis outside education was established in the 1930s for water resource projects and over time as been applied to a wide range of policy interventions; see Gramlich, 1981, and Boardman et al., 2011.) In terms of basic principles, the respective methods of CEA and CBA have not changed significantly in the intervening years. Yet education researchers have sparsely applied CBA and very sparsely applied CEA. In contrast, the literature on program impacts, which pays no attention to the cost of the programs, is large and methodologically more sophisticated. Thus, policymakers are gaining a better understanding of what works, but the evidence on what interventions, programs, and policies are comparatively most efficient is often not available.

In this chapter I set down the companion methods of CBA and CEA, beginning with a description of costs analysis. I then look at cost-effectiveness analysis, as this is conceptually less complex than CBA. I show how to integrate costs with measures of effectiveness to yield the key metric—the cost-effectiveness ratio. I then identify some important methodological challenges to performing CEA. Next, I conclude with a summary of current practice, along with some case studies as examples for future reference. I perform a parallel exercise for CBA with a discussion of metrics, methodological challenges, and evidence from policy areas where CBA has been used. In the final section I consider how education policy might be improved by more extensive use of CBA and CEA.

**Analysis of Costs**

Economic evaluations focus on the resources required to implement the program. The monetary value of these resources is referred to as the cost of the program. For both CBA and CEA, costs data should be collected using the ingredients method (Levin & McEwan, 2001; Tsang, 1997). This method requires the specification of all the ingredients or inputs that are used to implement the intervention net of the costs of the alternative or “business as usual” program. Thus, the costs are expressed as incremental beyond what is usually spent. Many education interventions change or supplement only a part of the total education a student receives: an after-school program, for example, adds extra schooling on top of the regular school day; a new reading curriculum affects only class time spent on reading. It is critical to estimate costs in relation to business as usual and with recognition that the program costs are incremental on top of other resources that improve educational outcomes.

All inputs should be costed out, regardless of who pays for them, but the financing source should be
identified. Resources for education are funded from many sources, including several levels of government and students’ families, as well as local agencies (businesses or community groups). Typically, the total cost of all these inputs is reported. However, costs for each perspective and funding agency (such as a school, district, or level of government) may be reported separately if needed. Conventionally, there are three perspectives. The primary perspective is the social, as this includes all resources. But education programs may also be evaluated from the perspective of the private individual participating in the program. Also, they may be evaluated from the fiscal perspective, that is, accounting only for the resources associated with government (and funded by the taxpayer). Results may differ depending on the perspective. For example, a school might contribute teacher time only to a program where the materials and facilities are funded by the federal government. From the school’s perspective, but not necessarily from a broader social perspective, this program might look very cost-effective.

Program ingredients should be grouped into personnel, materials/equipment, facilities, and other. As education is labor intensive, the largest ingredient in most reforms is likely to be personnel and especially teacher time. For some interventions, such as computer-based instruction or online learning, materials and equipment may be significant. Careful attention must be paid to amortizing this equipment over the life of the intervention: one claim made by providers of online learning, for example, is that the high initial investment lasts many years and reduces the costs of other inputs. For other educational programs, parent and student time commitments may be significant. Accurate accounting for student time is critical: many college and high school students may wish to work instead of enroll in supplemental education programs and so lost wages represent their opportunity cost of participation. Typically, collecting this information requires semistructured instruments administered to key personnel responsible for implementing the intervention (or observations of practice). These instruments should be framed to determine what ingredients are needed (not what amounts were spent). All these ingredients should be tabulated in a spreadsheet.

Once tabulated, each ingredient should be priced out using independent prices from appropriate datasets of agencies such as the Bureau of Labor Statistics or the National Center for Educational Statistics. Prices should capture what it would cost to purchase these ingredients in a perfectly competitive market. Many ingredients in education are purchased in distorted markets (e.g., ones with significant government regulations) and so attention must be paid to whether the prices used reflect their opportunity cost. For parent and student time, for example, it may be appropriate to use the market wage as the opportunity cost of time. But parents may enjoy helping out in schools and so require a lower “wage.” There is no consensus on how to adjust for this enjoyment: typically, the opportunity cost of parent time is accounted for by using a relatively low wage (e.g., the minimum wage) and applying this uniformly across all parents regardless of their occupation or employment status. All prices can be adjusted for inflation and local price indices applied as appropriate. The cost of the intervention is then reported as the total sum of all ingredients multiplied by their unit prices.

An example of the ingredients method is given in Table 9.1 for the federal Talent Search program (Bowden, 2013). Talent Search, initially established and funded through the 1965 Higher Education Act, is a program for high school students that includes tutoring, financial awareness and career selection training, college tours, and assistance with all aspects of applying to and enrolling in college. The inputs to the program include services provided by Talent Search offices and time and resources from the high schools and local community groups. Table 9.1 shows the inputs and the resources for each input aligned with the agency responsible for funding these resources. The final column shows the total resource requirement for Talent Search. The data are based on the inputs reported from interviews with five currently operating Talent Search sites. Information on the prices for personnel was taken from the National Compensation Survey (Bureau of Labor Statistics) and from the National Education Association salary survey. These prices (wages) were based on the respective occupations of the personnel, accounting for their experience and qualifications, and were also adjusted to account for fringe benefits where appropriate. Information on the prices for facilities was based on the rental rate for comparable office space. Information for other inputs was based on market prices for specific equipment and materials. The opportunity cost of time of other personnel (volunteers) was based on wage rates for workers with comparable skills and experience. As shown in Table 9.1, the largest proportion of the total cost is for personnel. Although the Talent Search agencies incur most of the costs of providing the program, there are significant resource commitments at the school site and from other groups (nonprofit agencies).

Table 9.1 Cost Spreadsheet Using the Ingredients Method
Costs analysis of Talent Search shows variation across sites in what resources are used. Other cost analyses have shown variation in how much resource is used. For example, a cost analysis of reading programs using the ingredients method is reported by Levin, Catlin, and Elson (2007). While the reading program was intended to be applied uniformly across sites, Levin et al. (2007) found substantial variation across sites in the total costs of implementation, with three times as much spent at the most resource-intensive site compared to the least intensive.

Crucially, cost estimates should not be derived from budgetary information. Agency budgets are rarely comprehensive, covering all relevant ingredients, and are almost never itemized in a way that clearly identifies how much is spent on a given intervention. Using budgets it is very difficult to disentangle the incremental costs of an education reform from the regular school or college operations. Further, budget statements reflect local prices and not what an intervention would cost if another agency decided to implement it in its local context. If budgets were used, interventions in areas with high prices would be disadvantaged in any economic comparisons.

Some interventions may induce additional resource use after the program is implemented. For example, a high school intervention to improve access to college will lead to increased spending for those students who now attend college. For clarity, it is better to separate out these “induced costs” (sometimes called “indirect costs”) from program delivery costs. For CEA, these induced costs may be included in the total costs (depending on whether they are necessary to effect the project goals). For CBA, these induced costs are often more accurately labeled as “negative benefits.” This is especially important when the results are reported as a benefit-cost ratio (see discussion below) because labeling them as costs will change the benefit-cost ratio by inflating the denominator.

Cost-Effectiveness Analysis

CEA is a very flexible method by which to supplement effectiveness analysis. It is possible to use any domain
of effectiveness for cost-effectiveness analysis. That is, one can perform CEA for interventions that improve phonics outcomes for struggling readers, high school graduation rates for seniors, or credits accumulated by college students. As long as the outcomes can be validly attributed to the intervention, they are legitimate measures of effectiveness for CEA.

However, effectiveness must be enumerated into a singular measure. If an intervention has multiple outcomes then these outcomes must be weighted to yield a singular measure of overall effectiveness. For example, an after-school program may improve students’ academic and social skills, and both outcomes should be included if they contribute to the outcome of interest such as graduation from the program. Weights may be determined by reference to theory or according to decision makers’ preferences. For example, educational programs may be funded based on how many students graduate from the program; that funding imperative will influence decision makers’ preferences. A more formal approach is to base the weights on an explicit cost-utility analysis (CUA). This type of analysis requires information on the value of each outcome. This value is based on preferences, which are typically elicited from education professionals or expert panels (or sometimes from parents). Once this valuation has been established, CUA follows the same method as CEA with utility substituting for effectiveness.

Cost-Effectiveness Ratios

In principle, the cost-effectiveness ratio is straightforward to calculate. It is an expression of the total costs of an intervention divided by the effects of that intervention. Its interpretation is also straightforward, again in principle. Interventions with the lowest cost per unit of effectiveness (smallest ratios) are the most cost-effective. In some cases, cost-effectiveness may be reported as the increase in effectiveness per dollar spent.

However, one must be very clear about how this ratio is actually derived. CEA is a comparative evaluation method. There must be at least two alternatives for comparison. At a basic level, the scenarios must be comparable in several important respects. As noted above, they must be intended to improve the same outcomes and these must be enumerated in a singular form. As well, the scenarios must be of similar scale, covering equivalent numbers of students. Large interventions cannot be easily compared to small interventions, because cost estimates cannot be easily extrapolated up or down. The example of a small-scale, teacher-led classroom intervention that is to be implemented statewide illustrates the three reasons extrapolation becomes difficult. First, the cost per student will increase: to recruit many extra teachers across the state, higher wages will have to be offered. Second, as the absolute cost of the intervention rises, it is likely that the state will hit a financing constraint or face a higher interest rate to borrow funds to implement the program. Third, the intervention may exhibit economies or diseconomies of scale: implemented statewide, the intervention may be more effective (through peer learning) or lower cost (with savings on administration and management). Input prices, financing constraints, and economies of scale may all be uncertain when extrapolating from a small to a large program.

The comparative aspect of CEA also has implications for the interpretation of the results. For a field trial outcomes for the treatment and control groups are compared. Both groups probably receive some educational resources (in regular classroom settings), but the treatment group receives more or different resources, for example through extra tutoring or in an after-school program. The costs of the intervention are therefore incremental costs, that is, they are extra resources beyond those all children (or all children in the control group) receive. The cost-effectiveness metric is therefore an incremental cost-effectiveness ratio (ICER). It expresses how the extra resources for the treatment generate gains in outcomes relative to the control group. With ICERs, the treatment is not “buying” outcomes but buying improvements in outcomes.

Expressed using these metrics, the value of cost-effectiveness analysis becomes clear. Interventions that are highly effective in the sense of generating big gains in valued outcomes may not be very cost-effective. That is, the reason why they are so effective might be because they require a large amount of resources. Instead, other interventions might be only slightly less effective yet sufficiently cheaper, and thus more cost efficient. CEA can help identify dominated interventions (those that are less effective and more costly) and dominating interventions (those that are more effective and less costly). However, a possible scenario is that interventions are like consumer products: the more expensive the intervention, the more effective it is. If so, the decision maker would need additional information to select an intervention using additional information, such as which
interventions are affordable.

**Empirical and Methodological Challenges**

Cost-effectiveness analysis has been applied in only the most limited fashion in education research. Many studies that claim to have performed such analysis have typically made only a cursory or partial attempt. In part, this situation has arisen because of, as noted by Harris (2009) and in writings by Henry Levin, the empirical and methodological challenges to performing CEA.

Obtaining cost data is challenging. Potentially, it requires access to many personnel, including those in charge of managing an intervention, those delivering it, and the students experiencing it. These personnel will not typically have collected the information needed on costs, and so multiple interview and survey instruments must be designed and administered. Cost estimates should be calculated during the intervention and not after the evaluation of effectiveness. When collected retrospectively, cost data are much less accurate: personnel cannot easily account for their time dedicated to an intervention; many of the personnel have taken on different roles; and typically there is limited archival data on how an intervention was implemented. Finally, as with effects, costs will vary across participants or sites. However, because costs are bounded at zero, this variation is unlikely to follow a normal distribution. The challenge is compounded by the fact that rarely do sampling frames for evaluations explicitly address variation in resource use. When such variation is discovered, it is then hard to estimate variation across the population. (Most likely, evaluation sites will have been chosen based on their power to detect statistically significant impacts not statistically significant cost-effectiveness.)

Integrating costs and effectiveness data is also challenging. This is the case even when effectiveness has been rigorously identified, for example through a field trial. Very few educational interventions have only one single outcome, so it may be necessary to apply weights or to apportion costs to particular outcomes (separating out the costs into those designed to improve each outcome). Yet there is little consensus on how to weight different effects. Effect size gains (differences in means divided by the standard deviation) are also difficult to interpret in terms of incremental cost-effectiveness ratios. These effect size gains depend on the variance in the populations being studied, so comparisons should only be made across interventions with the same variance in effects. Also, studies tend to report effects either for the intent-to-treat or treatment-on-the-treated samples. This choice has implications for how resources should be calculated. The intent-to-treat sample is often larger than those who actually receive the treatment and so the total costs will be higher.

Finally, cost-effectiveness results should be tested for robustness, that is, to see how ratios change with variations in estimates of costs or effects. Sensitivity analysis may be performed by selecting alternative values for the most important parameter or by reporting best-case and worst-case scenarios. Sensitivity analysis is necessary for several reasons. Many analyses are based on small samples with wide sampling error in estimates both of costs and impacts. Also, the variance in costs may not follow a normal distribution and so may preclude standard statistical tests for differences in means. Given variance in costs and variance in effects, the variance in cost-effectiveness may be difficult to predict. Potentially, the analysis may yield imprecise estimates of cost efficiency (particularly if the results are disaggregated to the site level), and this imprecision should be reported explicitly in a formal sensitivity analysis. As noted above, scale is one important consideration. Another is that input prices may change as technology changes (e.g., for school computing facilities).

**Examples and Findings**

There are few exemplary cost-effectiveness analyses in education (Levin & McEwan, 2002; Levin, 2001). In a review of over 1,300 relevant academic papers in education on cost-effectiveness, Clune (2002) divided them on a quality scale as follows: 56 percent “rhetorical,” 27 percent “minimal,” 15 percent “substantial,” and 2 percent “plausible.” As an indicator of the low quality of the research, Clune’s (2002) definition of “substantial” was an “attempt to mount data on cost and effectiveness but with serious flaws.” “Rhetorical” was defined as “cost-
effectiveness claims with no data on either costs or effects.” CEA of education policy has not improved considerably since Clune’s (2002) review. This paucity of evidence means that it is only possible to speculate which programs might be cost-effective (Harris, 2009).

One early example of CEA is the investigation by Henry Levin, Gene Glass, and Gail Meister (1987) of computer-assisted instruction. The authors compared the effectiveness of four interventions in improving mathematics and reading performance of elementary school children. These interventions were computer-assisted instruction; cross-age tutoring; reducing class size; and increasing instructional time. Each intervention had been found to be effective at improving achievement, but there was substantial variation in the costs of implementing each intervention. By comparing cost-effectiveness ratios, cross-age tutoring appeared significantly more efficient than computer-assisted instruction, which in turn was more efficient than either reducing class size or increasing instructional time. Peer-tutoring appeared most efficient for two reasons: the peers and the tutors both improve their achievement levels; and the time of the peers is valued at an opportunity cost of zero, as these students would be studying in class anyway.

Table 9.2 Cost-Effectiveness Results for Programs That Improve the Rate of High School Completion

<table>
<thead>
<tr>
<th>Program</th>
<th>(1) Treatment group size</th>
<th>(2) Treatment group graduation rate</th>
<th>(3) Control group graduation rate</th>
<th>(4) Percentage point gain (2) – (3)</th>
<th>(5) Yield of new graduates</th>
<th>(6) Cost per participant</th>
<th>(7) CE ratio (1)×(6)/(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention for students in school:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talent Search</td>
<td>3,930</td>
<td>82.5%</td>
<td>71.8%</td>
<td>10.8</td>
<td>423</td>
<td>$3,290</td>
<td>$30,520</td>
</tr>
<tr>
<td>Interventions for youth who have dropped out of school:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NGYC</td>
<td>596</td>
<td>87.0%</td>
<td>67.3%</td>
<td>19.8</td>
<td>118</td>
<td>$14,100</td>
<td>$71,220</td>
</tr>
<tr>
<td>Job Corps</td>
<td>3,940</td>
<td>51.4%</td>
<td>34.4%</td>
<td>17.0</td>
<td>670</td>
<td>$22,290</td>
<td>$131,140</td>
</tr>
<tr>
<td>JOBSTART</td>
<td>1,028</td>
<td>44.4%</td>
<td>29.3%</td>
<td>15.1</td>
<td>155</td>
<td>$10,460</td>
<td>$69,510</td>
</tr>
<tr>
<td>Intervention for young mothers who have dropped out of school:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Chance</td>
<td>1,240</td>
<td>58.6%</td>
<td>49.5%</td>
<td>9.2</td>
<td>113</td>
<td>$17,820</td>
<td>$194,640</td>
</tr>
</tbody>
</table>

Source: Hollands et al. (in press, Table 2).

Notes: Treatment on the treated (TOT) estimates for percentage point gain of treatment group over control group. 2010 dollars.

An example of a recently completed cost-effectiveness analysis is given in Table 9.2. This CEA is from Hollands et al. (in press), who reviewed all interventions that have been proven to increase the high school graduation rate and performed cost-effectiveness analyses on those for which both effectiveness and costs data could be obtained. The effectiveness measure was the incremental gain in high school graduates as a result of the intervention: columns 2–4 show that the interventions increased the high school graduation rate by 9–20 percentage points. Cost data were collected from archival sources and, for Talent Search, from a direct application of the ingredients method. Column 6 of Table 9.2 shows the large differences in the cost per participant for each intervention, ranging from $3,290 to $22,290. The cost-effectiveness ratios are reported in column 7. The CE ratio is the total cost across the entire sample (entries in column 1 times those in column 6) divided by the yield of new graduates (column 5). The cost for the entire sample must be used, even as some of these students would have graduated anyway, and some do not graduate even after the intervention. This CE ratio can be interpreted as the incremental cost to yield a new high school graduate. Lower ratios are therefore preferred as these represent lower costs for ensuring a student graduates from high school. There are sizeable differences in the cost-effectiveness of these interventions: Talent Search is the most cost-effective, with the other interventions being less than half as efficient. Per dollar, investments in Talent Search will yield much greater improvements in the high school graduation rate than investments in other interventions. However, as noted above, a critical factor in interpreting cost-effectiveness ratios is the context of the decision. The interventions listed in Table 9.2 serve very different student groups. Talent Search is an intervention for students who are in school; the other interventions are for students who have already dropped out of school. Getting students who are in high school to complete their studies is much less costly than getting dropouts to complete their high school studies. Nevertheless, cost-effectiveness analysis provides information on the extent of the disparity in resources required to ensure high school graduation for different subgroups.
Cost-Benefit Analysis

Cost-benefit analysis can potentially be performed for any educational intervention. Using the ingredients method, the costs of each intervention can be expressed in dollars. For CBA, the impacts and outcomes of the intervention are also expressed in monetary terms and labeled as the benefits. The simple logic of CBA is that policies are preferred where the benefits exceed the costs. The advantage of CBA over CEA is that each outcome of an educational intervention can be explicitly accounted for in that many outcomes over many periods of time can be considered. Consequently, interventions with different goals and different outcomes can be compared, including interventions unrelated to education. For example, the returns from an education program may be compared to those from a public infrastructure project. However, the validity of CBA depends on the extent to which outcomes can be expressed in dollars. If many outcomes cannot be monetized, then CBA may not be an appropriate way to evaluate an intervention, although even in this case it may be helpful to use the framework of CBA to systematically organize the analysis and identify the key unknowns.

Most educational outcomes—such as test score gains or increased rates of college completion—do not have clear market prices. They cannot be “bought,” but they are valuable; they lead to a more productive citizenry. Hence, it makes economic sense for society to commit resources to obtain these outcomes. To translate these educational outcomes into dollar values, indirect methods, often called shadow pricing techniques, are needed. A shadow price is the amount of resource someone would be willing to pay in order to obtain a particular outcome.

There are several ways to calculate shadow prices and they may be calculated from the same three perspectives as used above for costs: individual, fiscal, and social. The most common shadow pricing approach is to add up all the economic benefits that flow from reaching a particular education level (or add up all the burdens from not being educated). This method is akin to the “cost of illness” method in the health economics literature (and might be called the “cost of ignorance” method in the education sector). For example, from the individual’s perspective the shadow price of being a high school graduate as opposed to a high school dropout is the additional lifetime earnings associated with graduating from high school. From the fiscal perspective, the shadow price might be additional tax payments from being a high school graduate. This shadow price is conservative even as a measure of fiscal savings because it does not include the value of savings to the criminal justice, health, and welfare systems that arise from having a more educated populace (see Belfield & Levin, 2007). It is obviously even more limited as a measure of social benefits.

As another example, the shadow price of reducing the rate of special education may be calculated as the savings that arise from placing a student in mainstream instruction rather than special education classes. An alternative shadow pricing technique is to derive the economic value of education from the prices of other goods that achieve the same outcomes. For example, the shadow price of a high quality education can be calculated as the extra amount one has to pay for a house that is close to a high quality school. The shadow price for a higher test score might be calculated as the price paid for private tutoring that is needed to generate that test score gain. There are existing shadow prices for some educational outcomes that might be “plugged in” to an impact evaluation. For example, Belfield and Levin (2007) estimate the fiscal shadow price per additional high school graduate at $209,100 in 2005 dollars (see Karoly, 2012). But for most outcomes, the researcher must calculate these prices independently, and even when shadow prices exist, care must be used in transferring them to other contexts.

Shadow prices for educational outcomes must be adjusted to account for when these outcomes occur. So, a program in 10th grade that increases high school graduation rates (impacts) should lead to higher earnings (benefits) in adulthood. But these benefits lag behind the costs: the costs of the program were incurred in 10th grade and the flow of benefits will not occur for at least two years. The benefit flow must therefore be discounted on the grounds that money received in the future is worth less than money received in the present (see Moore et al., 2004). The practice of discounting involves progressively lowering the value of dollar amounts as they occur further into the future (and follows the practice illustrated in accounting textbooks). This weighting is reflected in the choice of the discount rate. Choosing a high discount rate means future benefits will be worth less than if a low discount rate were chosen. The value of the discount rate may vary depending on the types of benefits that accrue, but typically values of 3–7 percent are chosen for educational interventions (provided the benefits are reported in constant dollars—they have been deflated for price
increases). When benefits and, if necessary, costs have been discounted, they are referred to as present values; these are the value figures that should be compared for the CBA.

**Cost-Benefit Metrics**

The comparison of costs and benefits can be expressed using three different metrics. The first is benefits minus costs; as these are expressed in present values, this difference is called the net present value (NPV) of the educational investment. Interventions where the net present value is positive are worth undertaking, but given competing interventions the one with the largest net present value is preferred. The second metric is the benefit-cost (B/C) ratio, which is the benefits divided by the costs. If this B/C ratio exceeds one, the benefits are greater than the costs and the investment is worth undertaking. Typically, the intervention with the highest B/C ratio is preferred. The B/C ratio is quoted most often in CBAs because it is a useful summary of the economic value (as in “for each dollar spent on the program, $X dollars are returned in benefits”). Importantly, and unlike the net present value, the B/C ratio is independent of scale and so must be interpreted carefully. An education program that costs $1 million but yields benefits of $3 million has a B/C ratio of 3 and a net present value of $2 million. By comparison an educational intervention that costs $1,000 but yields benefits of $10,000 has a B/C ratio of 10 and a net present value of $9,000. In monetary terms, the former intervention would be preferred even as it has a lower B/C ratio.

The final metric is the internal rate of return. This is the discount rate that ensures that the present value of benefits equals the present value of costs. Instead of assuming a discount rate (of say 3 percent), calculating the present values of costs and benefits, and reporting the difference, the researcher derives a value for the discount rate such that the difference between the costs and benefits is equal to zero. This value of the discount rate is called the internal rate of return (IRR). Again, educational interventions with high IRRs are worth undertaking, and the one with the highest IRR is preferred. For this metric, the IRR might be compared to an outside standard such as the rate of return on Treasury bonds or shares. So, if the government can borrow money at 4 percent and provide a reading program that yields a 10 percent return (IRR), then the program is a worthwhile investment.

Each of these metrics can be calculated from each perspective. For the private individual, most education interventions easily pass a cost-benefit test because the student typically pays only a small proportion of the costs but reaps a large proportion of the benefits (e.g. higher earnings). From the fiscal perspective, the calculus is less clear-cut; taxpayers often pay a large proportion of the costs but reap a relatively small proportion of the benefits (e.g. increased tax revenues on higher earnings). For society, the full resource costs must be compared against the full social benefits.

**Empirical and Methodological Challenges**

As with CEA, there are both empirical and methodological challenges to overcome in conducting CBA. As CBA applies the same method of costing as CEA, the same challenges as noted earlier apply in obtaining costs data and linking it to the impacts of the intervention.

However, different challenges arise in estimating benefits because of the difficulty in accurately estimating shadow prices for educational outcomes. Even if the outcomes of an educational intervention are identified (e.g., from a random assignment trial), the economic value of these outcomes may not be precisely bounded. For example, the effect size gain in achievement from the Tennessee STAR experiment to reduce class size may be precisely identified (see, for example, Finn et al., 2005), but there is much less certainty over the economic value of reducing class size.

Broadly, educational outcomes can be grouped into those related to attainment (e.g., years of schooling or college), achievement or cognitive test scores, and behavioral change (e.g., reduced delinquency or increased socioemotional knowledge). Presently, there are well-established protocols for estimating shadow prices for the first of these groups. That is, there is reasonable evidence on the value in terms of higher incomes from getting students to graduate from high school and from each additional year of education (at least the income effects). However, there is much less information for the other two groupings. Notably, despite substantial policy
attention to raising test scores and reducing achievement gaps (e.g., No Child Left Behind), there is no consensus on the economic value of achievement gains. In order to identify these gains, it is necessary to separate out confounding factors—students with higher achievement tend to have higher attainment—and to specify how initial cognitive gains influence subsequent cognitive skills (for a discussion, see Belfield & Levin, 2009). For delinquency, there are shadow prices for specific crimes (e.g., violent assault) and there are some estimates of the cost consequences of being an at-risk youth (Cohen & Piquero, 2009). However, these prices are averages across all youth and it may be that educational investments are influential only for intermittent offenders and not chronic offenders. Few shadow prices for socio-emotional skills have been calculated.

As with CEA, the response to these challenges is to perform extensive sensitivity testing. Alternative values for key parameters and best-case/worst-case sensitivity tests may be performed. Indeed, worst-case sensitivity analysis may be sufficient to substantiate the results of a CBA: if the benefits still exceed the costs with a worst-case scenario, then it is very likely that the program is allocatively efficient. Similarly, if a CBA includes only a few of the benefits of an education program, and yet these nonetheless exceed the costs, then this too generates a strong conclusion about efficiency.

In addition to these sensitivity tests, for CBA Monte Carlo simulations may be applied. These simulations use information on the distributions of costs, impacts, and shadow prices. By repeatedly drawing new estimates from the distributions of values for each parameter, the NPV, B/C ratio, or IRR can be re-derived. (Typically, 500 or 1,000 draws are taken.) Monte Carlo simulations are helpful when the impacts of some interventions are not statistically significant from the comparison group. The wide variation in these impacts will be reflected in the variance in the economic metrics. These simulations provide information not only on the precision of the net present value but also on the probability that this NPV is negative (i.e., that the costs exceed the benefits).

**Examples and Findings**

Evidence on the costs and benefits of educational programs is more readily available than evidence on their cost-effectiveness (Hummel-Rossi & Ashdown, 2002). Much of the available CBAs are for early childhood programs, although there are now more examples of CBA at the high school and college levels.

The Chicago Child-Parent Center program is one example of a preschool program that has been subjected to a full CBA. (Others include the High Scope/Perry Preschool program and the Abecedarian Child Development Program; see Heckman et al., 2010, and Barnett & Masse, 2007.) The preschool program is intended to help promote children’s academic and social skills with the intention that this will enhance the children’s development, leading to improved economic well-being in the short, medium, and long run. Specifically, investments in preschool are expected to: help families with child care responsibilities; improve resource use in school (e.g., by reducing the need for special educational services); and, by enhancing the child’s human and social capital, lead to higher incomes and lower delinquency. These consequences can all be expressed in monetary terms.

The costs and benefits of the Chicago Child-Parent Center program are shown in Table 9.3 (from Reynolds et al., 2011). These are based on follow-up of Center preschool participants at age 26 compared to follow-up of a quasi-experimentally matched control group. Three perspectives are reported: those of the child participant, the general public (those other than the participants), and society (the sum of the child and general public impacts). The preschool program costs $8,512 per student, which is paid by the general public and not the child’s family. The benefits of preschool are substantial and diverse across the different perspectives. For the individual, the benefits include: childcare savings for the family, improved child welfare (lower abuse/neglect), and significantly higher earnings over the lifetime as the child becomes an adult. In total, the present value benefits to the child are $30,974, and these are also the net benefits because the program is at zero cost to their families. The general public perspective shows there are many other benefits from preschool. These include: improved child welfare, savings to education budgets, increased earnings, and reductions in government spending on criminal behavior and health. The total present value benefits to the general public are $61,246. The net present value is therefore $52,734 ($61,246 minus $8,512) and the benefit-cost ratio is 7.2:1. From the social perspective, the benefits measured across all these domains are substantial at $92,220; the net present value is $83,708, and the benefit-cost ratio is 10.8:1. Finally, from the social perspective, the internal rate of
return is 18 percent. Using any of these metrics, the Chicago Child-Parent Center preschool program easily passes a cost-benefit test. Also, the more intensive is the program, the greater are the net benefits. Programs of one year or less generated net benefit amounts of $68,347 versus $83,708 for programs lasting more than one year (Reynolds et al., 2011, Table 5).

Table 9.3 Chicago Child-Parent Center Program (Pre-School): Cost-Benefit Analysis

<table>
<thead>
<tr>
<th>Perspective</th>
<th>Participants</th>
<th>General Public</th>
<th>Society</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program Costs [C]</td>
<td>–</td>
<td>$8,512</td>
<td>$8,512</td>
</tr>
<tr>
<td>Program Benefits:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child care</td>
<td>$4,387</td>
<td>–</td>
<td>$4,387</td>
</tr>
<tr>
<td>Child abuse/neglect</td>
<td>$4,240</td>
<td>$3,090</td>
<td>$7,330</td>
</tr>
<tr>
<td>Education</td>
<td>$(98)</td>
<td>$6,001</td>
<td>$5,903</td>
</tr>
<tr>
<td>Earnings</td>
<td>$22,445</td>
<td>$6,399</td>
<td>$28,844</td>
</tr>
<tr>
<td>Criminal behavior</td>
<td>–</td>
<td>$42,462</td>
<td>$42,462</td>
</tr>
<tr>
<td>Health</td>
<td>–</td>
<td>$3,294</td>
<td>$3,294</td>
</tr>
<tr>
<td>Total Benefits [B]</td>
<td>$30,974</td>
<td>$61,246</td>
<td>$92,220</td>
</tr>
<tr>
<td>Net Present Value (Benefits-Cost)</td>
<td>$30,974</td>
<td>$52,734</td>
<td>$83,708</td>
</tr>
<tr>
<td>Benefit-Cost Ratio</td>
<td>–</td>
<td>7.2</td>
<td>10.8</td>
</tr>
<tr>
<td>IRR</td>
<td>18%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Reynolds et al. (2011, Table 4, Appendix C). 2007 dollars. Discount rate 3 percent. Education benefits are savings in grade retention and special education net of additional spending on college. Health benefits are for reduced depression and substance abuse.

CBAs of early childhood interventions have been summarized by Karoly (2012), and the results for social CBAs are presented in Table 9.4. These results show that for each program the benefits substantially exceed the costs, in some cases by a factor of more than 10. A general finding of the CBA literature is that early interventions yield high net present values and IRRs.

Table 9.4 Social Benefit-Cost Ratios: Early Childhood Interventions

<table>
<thead>
<tr>
<th>Program</th>
<th>Benefit-Cost Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurse-Family Partnership</td>
<td>2.88</td>
</tr>
<tr>
<td>Abecedarian Child Intervention</td>
<td>2.49</td>
</tr>
<tr>
<td>HIPPY USA</td>
<td>1.80</td>
</tr>
<tr>
<td>Chicago CPC</td>
<td>10.83</td>
</tr>
<tr>
<td>Perry Preschool</td>
<td>7.1–12.2</td>
</tr>
</tbody>
</table>

Source: Adapted from Karoly (2012, Table 2). Discount rate 3 percent.

Conclusions

Ultimately, the value of CEA and CBA rests on the extent to which it improves policy decision making. Purposefully, cost-effectiveness analysis and cost-benefit analysis are intended to help decision makers allocate resources to educational programs as efficiently as possible. The conditions under which CEA is appropriate are strict and the data requirements for CBA are quite challenging. Nevertheless, even an analysis of costs can be helpful in clarifying how an intervention is actually delivered on the ground and by showing how much it costs when all resources are accounted for. It is possible that some reforms fail to be effective because they do not involve much change (e.g., curricular reforms). It is also possible that some reforms cannot be implemented because they are just too expensive (e.g., reducing class size or whole school reform) or rely on volunteer labor
that is not available in large scale. Costs analysis is informative on these possibilities. Even where CEA and CBA are imprecise, they may still provide information on programs that clearly do not pass an efficiency test or on ones (such as early education) that clearly do.

For CEA and CBA to be more widely useful, policymakers need to be able to draw general conclusions from the available evidence. That is, they will want to compare the returns from one investment to the returns from another so as to make the most efficient choice. Therefore, it is necessary for the evidence to be expressed in metrics that are comparable. Indeed, this is a significant advantage of CBA: because the evidence is expressed in monetary units, then potentially any intervention can be compared against another. Yet, if analyses are not harmonized there is a danger that policymakers will draw false inferences about efficiency. Of course, as there are so few CEAs and CBAs, policymakers might also be concerned about the many interventions whose efficiency has not yet been investigated.

There are many dimensions on which CEA and CBA studies need to be harmonized (Karoly, 2012). These dimensions include the range of outcomes considered—a study will have a higher benefit-cost ratio simply if it counts more benefits (e.g., if both labor market earnings and health status gains from high school graduation are counted instead of just the former). Karoly (2012, Table 6) catalogs the many differences in outcomes evaluated in CBAs of early childhood interventions. Other important dimensions are the time horizon of the evaluation—looking over a longer time horizon will likely yield a higher estimate of benefits—and the discount rate for deriving present values. On the latter, Karoly (2012) proposes a 3 percent discount rate should be the standard, with sensitivity analysis performed using discount rates of 0, 5, 7, and 10 percent. More generally, most CEAs and CBAs do not include any distributional considerations. An intervention to help struggling students in college is evaluated using the same tools and metrics as an intervention to help struggling readers in kindergarten. The units (dollar costs, impacts, and dollar benefits) do not differ across these two interventions. Although this equivalence is explicit in CEA/CBA, it is likely that decision makers will find it hard to adjudicate between programs serving very different student groups.

Fundamentally, it would be useful for more CEAs and CBAs to be performed within the education sector. The large aggregate resource commitment—both in money and the time of K–12 and college students—makes it important that decisions about educational investments are made efficiently. CEA/CBA, as well as basic analysis of costs, can contribute toward improving the quality of decision making across the sector.

References


Teachers Clearly Matter, But Finding Effective Teacher Policies Has Proven Challenging

The Value of Teacher Quality

For most it comes as no surprise that research has consistently shown that teachers can have profound effects on their students. Research dating all the way back to the “Coleman Report” (Coleman, 1966) found teacher quality to be the most important schooling factor influencing student achievement. Recent findings, based on better data and more rigorous statistical methods, confirm this conclusion and furthermore show that there is considerable variation in quality among teachers (Aaronson et al., 2007; Goldhaber et al., 1999; Rivkin et al., 2005). Hanushek (1992), for instance, finds that the quality of a teacher can make the difference of a full year’s learning growth, and Chetty et al. (2013) find that differences in teacher quality explain a variety of later life outcomes, such as college attendance and labor market earnings.

While it is common to discuss teacher quality in abstract terms, the nuts-and-bolts definition of the concept is to a large extent in the eye of the beholder. One notable change in teacher research and policy over the last decade is the shift toward assessing teacher quality based on teachers’ “value added,” that is, their contributions to student achievement on standardized tests. Indeed the research cited above referring to consequential variation in teacher quality is based on value added. The use of value added to assess individual teachers is controversial for a variety of reasons (discussed more fully below). Probably the most fundamental reason is that student achievement on tests captures only one aspect of what teachers contribute to students. Teachers no doubt contribute in many different ways to students’ intellectual and emotional growth, implying that their quality should be measured along a number of dimensions. Historical measures of teacher quality have generally been associated with the set of credentials that a teacher holds. It is also commonplace to associate the quality of teachers with their classroom practices (e.g., Bell et al., 2012; Hill et al., 2011 and 2012). Nonetheless, in this chapter I primarily focus on research that uses gains in student test scores as a metric for teacher quality. I do this not only because test scores are arguably the best indicator of teacher performance (at least for the relatively small proportion of teachers for whom they can be calculated), but also for a more practical reason: a good deal of education research focuses on test scores as an outcome because there is little variation in other measures of teacher performance (Weisburg et al., 2009). Hence, throughout the chapter, I use the term “teacher quality” interchangeably with “teacher effectiveness,” “teacher performance,” and “value added.” For a richer discussion of teacher effectiveness, see Donaldson and Papay in this volume.

A large body of research suggests that while teacher quality (as just defined) is an important predictor of student success, it is not strongly associated with readily measurable teacher characteristics such as degree and experience levels. This point is well-illustrated by Goldhaber et al. (1999), who first estimated the overall impact of teachers on students, then parsed out which part could be explained by specific teacher characteristics and which part was attributable to less easily quantifiable teacher attributes (e.g., whether they are enthusiastic and hardworking). They find that teachers account for approximately 8.5 percent of the variation in students’ 10th-grade achievement. The impact of unobservable teacher attributes (i.e., those that are not included in the dataset) dwarf that of objective, observable, easily measured attributes such as degree and experience levels: less than 5 percent of the variation in teacher quality is explained by such quantifiable characteristics. In the decade and a half since that research was completed, the basic finding that readily quantifiable teacher credentials and characteristics explain relatively little of the variation in teacher
effectiveness has been replicated many times. Yet the fact that teacher credentials explain little of the variation in effectiveness need not imply there is no relationship between credentials and effectiveness, and scholars reach differing conclusions about the degree to which particular credentials are educationally meaningful predictors of teacher quality. While a nuanced reading of the research literature suggests that some readily identifiable characteristics do predict success in the classroom such as early career teacher experience in particular (Rice, 2013), an important theme emerging from the more recent educational production function studies is that, when it comes to teachers, there is far more variation in the effectiveness of teachers who have a particular attribute (credential or characteristic, such as experience) than there is between teachers with different attributes. In other words, even when a particular teacher attribute does predict teacher quality in a statistical sense, there are greater differences between teachers who share a common attribute than there are differences between teachers with different attributes. In this chapter, I briefly explore the research linking teacher attributes and in-service measures of their performance to student achievement, and describe some of the policy implications related to this research.

Teacher Quality and Readily Observable Teacher Characteristics

The proliferation of studies made possible by the No Child Left Behind law’s annual testing requirements, combined with new administrative databases that link students to their teachers and track both over time, have permitted a significant amount of research on teacher attributes—licensure status, experience, degree level, and certification by the National Board for Professional Teaching Standards—that are particularly important to consider as they determine employment eligibility and compensation in the overwhelming majority of localities. In addition to discussing the relationship between these teacher attributes and student achievement, I also review what we know about the relationship between student achievement and measures of teachers’ academic competence. This relationship has been the subject of much policy discussion given comparisons between the teacher workforce in the United States and those in other countries with high-achieving students.

Licensure Status

The first step in becoming a teacher is getting approval to participate in the teacher labor market by the state in which you want to teach. The primary purpose of states’ regulation of employment eligibility, through licensure, is to assure the public that individuals in the teaching profession have at least the minimal standard of teaching competence to be qualified to begin practicing in the profession.

Teacher licensure systems typically require that prospective teachers complete a standard set of college-level courses in pedagogy and/or in the subject they wish to teach, and to demonstrate competence by passing one or more standardized tests. The last decade has also seen tremendous growth in the proportion of new teachers entering the labor market through alternative routes; some states rely on alternatives to traditional teacher training preparation quite heavily now (Boyd et al., 2007; Goldhaber, 2010). Earlier debates about the extent to which a teacher’s licensure status (also commonly referred to as “certification”) predicts teacher effectiveness were highly contentious. A number of recent high-quality studies shed new light on the issue.

Several large-scale experimental studies compare the effectiveness of teachers recruited into the profession through the alternative pathway of Teach For America (TFA) with other teachers in the same schools. What makes these studies unique is that the experiment involves randomly assigning students to TFA and non-TFA teachers in order to avoid the potentially confounding problems associated with matching students and teachers based on either observable or unobserved student characteristics. The studies find that, at the elementary level (Glazerman et al., 2006), the performance of students assigned to TFA teachers was comparable to non-TFA teachers in reading and exceeded that of non-TFA teachers in math; at the secondary level (Clark et al., 2013), TFA teachers outperformed non-TFA teachers in terms of students’ math performance (there was no comparison for other subjects).

Teach For America is one of the most visible alternative programs, but it is not very representative of most alternative routes into the classroom; TFA teachers tend, for instance, to have relatively high college entrance
exam scores (Decker et al., 2004; Rock-off et al., 2007). Another randomized control trial (Constantine et al., 2009) that focuses on less selective alternative routes to the classroom also finds little difference in the test achievement of students whose teachers received traditional training and those who entered teaching through alternative routes. It also shows little relationship between the amount and content of coursework that teachers completed while in training and student achievement. Again, this finding is reflected in much of the nonexperimental literature, which tends to show that the differences between teachers entering the profession through a specific route far exceed the average differences that exist between teachers entering through different routes (Boyd et al., 2009; Rockoff et al., 2007).

At the same time, these studies of alternative routes into the classroom may not tell us much about the value of teacher training itself because of both selection and training effects; in other words, the effectiveness of teachers will depend both on who enters a particular route and the specific training associated with that route. Given this limitation, it is not surprising that we know relatively little about the link between specific approaches to teacher preparation and the effectiveness of teachers in the field. In addition, research is limited on other potentially important ways that licensure could affect the quality, and distribution of quality, of teachers in the workforce, the choice of individuals to pursue a career in teaching, teachers’ cross-state mobility, and the hiring decisions made by local school systems (Goldhaber, 2010).

### Years of Teaching Experience

Teacher experience, and its value as a predictor of teacher effectiveness in the classroom, has been studied extensively, and while the precise magnitude of the estimated teacher-experience effect differs from one study to the next, most studies suggest that a teacher tends to become significantly more effective in each of the first three to five years of his or her career. The positive findings on teacher experience are found in different settings and span different grade levels and subjects, and thus appear to be robust to different teaching contexts (Boyd et al., 2008; Clotfelter et al., 2010; Goldhaber, 2007; Harris & Sass, 2011; Rivkin et al., 2005; Rockoff, 2004). Rivkin et al. (2005), for instance, estimate that elementary and middle school students of teachers with one year of experience are about 3 to 5 percent of a standard deviation better off in both math and reading than those with teachers who have no prior experience. In contrast, the gain from the increased experience achieved by going from the first to the second year of teaching is smaller (and borderline statistically significant) at around 2 to 3 percent of a standard deviation of student achievement. Rivkin et al. find little gain from experience beyond year two of teaching. Clotfelter et al. (2006 and 2010) find similar results. In math they find a gain between the first and second year of teaching of between 5 to 7 percent of a standard deviation, and in reading the gain ranges from 2 to 4 percent of a standard deviation. Some more recent research shows, however, that the returns to experience extend well beyond the first five years of teaching (Papay & Kraft, 2011; Ladd & Sorensen, 2014). For a review of the research on teacher experience, see Rice (2013).

The evidence that teachers tend to become more productive early in their careers is quite consistent, but this tendency masks the fact that there is a good deal of heterogeneity in the individual returns to experience (Atteberry et al., 2013). The majority of teachers do become more effective early on, but some do not, and some actually become less productive with experience. New research is beginning to explore whether the professional environment in schools might help explain why some teachers improve with experience while others do not (Kraft & Papay, 2013). This is a promising area for further investigation.

### Degree Level and Subject-Matter Preparation

There is relatively little evidence that teacher degree level, e.g., having a master’s degree, is generally a good predictor of teacher quality. While perhaps surprising given evidence from the broader economy of an association between years of education and productivity (e.g., Card, 1999), this finding is well documented by education research (Aaronson et al., 2007; Clotfelter et al., 2010; Goldhaber & Brewer, 1997; Harris & Sass, 2011; Podgursky & Springer, 2007; Rivkin et al., 2005) and in fact is not new (e.g., Hanushek, 1986). A plausible reason for the lack of a finding on teacher degree level is that many teachers get advanced degrees in areas that are not directly related to their teaching assignments, likely because school systems generally reward them
with a salary bump regardless of the specific area of the master’s program (Miller & Roza, 2012). There is, however, some evidence that a teacher’s degree level may matter in some specific contexts, especially in math at the secondary level, implying that subject-matter preparation is important. For example, Monk and King (1994) find that each additional course a teacher has taken in math improves student mathematics achievement by about three-quarters of 1 percent of a standard deviation, and similar science preparation increases student achievement in science by up to two-thirds of 1 percent of a standard deviation, findings that are similar in magnitude to those found by Goldhaber and Brewer (1997).

**NBPTS Certification**

The empirical evidence on the relationship between teachers holding a credential from the National Board for Professional Teaching Standards (NBPTS) and student achievement is more consistently positive than that described above for degree level, but it is still somewhat mixed and dependent on the nature of the comparisons. In particular, several studies (e.g., Cavalluzzo, 2004; Goldhaber & Anthony, 2007; Vandevoort et al., 2004) comparing the effectiveness of NBPTS to non-NBPTS certified teachers find NBPTS certification to be a statistically significant indicator of teacher quality. These studies all find positive effects of having an NBPTS-certified teacher in both math and reading test scores, though the magnitudes of the estimated effects differ. But not all the research on NBPTS teachers shows such positive effects (Cantrell et al., 2008; Harris & Sass, 2011; Sanders, Alston, & Wright, 2005). The differences in findings between NBPTS studies may be explained by differences in datasets, methodology employed, setting (e.g., much of the research on NBPTS takes place in Florida or North Carolina, given the large number of NBPTS certified teachers in those states), or the cohort of teachers being assessed.

More consistent is the evidence on whether NBPTS does a good job of identifying the more effective applicants. In particular, three (Cantrell et al., 2008; Cavalluzzo, 2004; and Goldhaber & Anthony, 2007) of the four studies that compare teachers who apply for NBPTS certification find that successful applicants are more effective than unsuccessful applicants. At the same time studies generally do not find that NBPTS teachers are more effective as a consequence of having gone through the NBPTS certification process (Goldhaber & Anthony, 2007; Harris & Sass, 2011).

**Teacher Test Performance and Demonstrated Academic Proficiency**

The last decade has seen an increased focus on the academic proficiency (or competence) of the nation’s teachers (Goldhaber & Walch, 2014), driven at least in part by high-profile reports comparing teachers in the United States to those in countries whose students score well on international test comparisons (Auguste et al., 2010). Academic proficiency is usually judged by measures such as teacher performance on verbal ability tests, teacher licensure exams, college entrance exams, or the selectivity of the undergraduate institutions attended by teachers. Although studies that include measures of teacher academic proficiency are far fewer than those that include degree and experience levels, the early research typically shows a positive relationship between individual teacher test scores and student achievement (Goldhaber, 2004; Greenwald et al., 1996). Yet the research on the effects of specific measures of academic competence are mixed. For instance, Harris and Sass (2011) find no relationship between teachers’ college entrance exams (SAT) and student achievement, whereas Clotfelter et al. (2010) find that teachers who attend more selective undergraduate institutions tend to be more effective. And in the studies of TFA described above, the differences in effectiveness between TFA and non-TFA teachers (in math) do not appear to be attributable to the fact that TFA teachers tend to have graduated from highly selective undergraduate institutions.

The evidence on licensure exams is more consistent. Several studies that explore the relationship between teacher performance on licensure exams and student achievement find a significant albeit modest positive relationship between the two (Clotfelter et al., 2006; Goldhaber, 2007). But, not all licensure exams—and there are a great many of them—have equal (or necessarily any) predictive power. In cases where the exams do predict student achievement, the difference in effectiveness between teachers who pass and fail licensure tests is relatively small (about 4 to 6 percent of a standard deviation of student achievement at the elementary level).
Importantly, while the tests are generally just used in an up or down way to determine employment eligibility, they do appear to provide information about teacher effectiveness beyond the simple pass/fail signal as there is, on some tests, a relationship between a teacher’s performance on licensure tests and student achievement even for teachers whose scores are far from the employment eligibility cutoff point.

More Variation within than Between and Teacher Gateway Policies

These findings suggest that a teacher’s readily observable characteristics are not generally strong indicators of teacher quality. This makes crafting policies designed to ensure a basic level of teacher quality based on pre-service credentials difficult: simply put, there do not appear to be readily quantifiable paper credentials that strongly predict teacher quality. Several recent studies aptly illustrate this conclusion. The Goldhaber (2007) paper on teacher licensure tests shows that there is far more variation between teachers who either pass or fail the North Carolina licensure test than there is between all those who pass and all those who fail. Kane et al. (2008) investigate the relationship between a teacher’s route into the classroom (e.g., traditional certification, TFA, Teaching Fellows Program) and student achievement, and find that there is much more variation in teacher quality within a route to the classroom than there is between routes; specifically, they estimate that the value-added differential in teacher effectiveness within a licensure category (i.e., having a teacher in the top quartile of effectiveness versus a bottom quartile teacher) is at least 10 times larger than the average gap between licensure categories. Put another way, even when a particular teacher attribute meets the standard of being "statistically significant," there is by no means a guarantee that a teacher with the attribute is going to be more effective than one without: there are, for instance, many fifth-year teachers who are less effective than the average novice teacher and vice versa, even if teacher experience tends to be a significant predictor of student achievement.

A potential policy implication of the finding that paper credentials do not accurately predict teacher quality is that policymakers ought to relax the certification gateway, allowing localities greater leeway over whom they select into the workforce. Then they ought to put more emphasis on rewarding and keeping those teachers in the classroom who are deemed successful, and weeding the unsuccessful teachers out of the profession. This policy prescription, which has been made by a number of researchers (e.g., Ballou & Podgursky, 1998; Gordon, Kane, & Staiger, 2006; Hanushek & Rivkin, 2003), is consistent with states’ growing reliance over the last decade on alternative certification programs that permit individuals with divergent pre-service experiences to move into the teaching profession, at least briefly (many programs require teachers to obtain “full” state certification within a few years of teaching).

While a relaxation of the certification gateway may seem a sensible shift in policy emphasis in light of the research findings described here, changes in pre-service requirements have ambiguous impacts on the quality and distribution of teachers in the workforce (Boyd et al., 2007; Goldhaber, 2004). For example, although relaxing the certification gateway might ultimately improve the average quality of the teacher workforce, it may allow more very low-quality teachers into the workforce, at least for a time. Whether this is a worthwhile trade-off depends on how one values the potential harm done by having more low-quality teachers in the workforce as compared to the potential benefit of having high-quality teachers in the workforce. Furthermore, after certification, local school systems act as a second gateway in determining the composition of the teacher workforce; therefore, how good they are in selecting teachers is an important mediating factor in determining the quality of the teacher workforce (Goldhaber, 2004). In theory at least, local hiring officials have the capacity to garner extensive information about teacher candidates—information that may go a long way toward explaining the significant variation researchers observe for teachers who are in the labor force.

Contrary to much conventional wisdom, many schools and districts have a large degree of choice when it comes to hiring teachers. For example, Ingersoll and Perda (2010) report that in 2000 (when the teacher labor market was relatively tight), the ratio of all teachers in the supply pipeline to the number of teachers leaving through retirement and attrition was over two to one. Similarly, Strauss et al. (2000) and Goldhaber et al. (2014) find there are far more prospective teachers credentialed than available teaching slots in general. But, this general finding masks the fact that there are usually far more applicants per slot in elementary education than in, for instance, the areas of science, technology, engineering, and mathematics (STEM) and special education.
Moreover, there is considerable heterogeneity across school districts in hiring options: high-poverty districts, for instance, are often at a disadvantage when it comes to filling teaching slots.

A relatively thin literature focuses on what schools or school systems look for when hiring teachers, and little information exists about the efficacy of specific processes schools (or districts) use in selecting among teacher applicants (DeArmond & Goldhaber, 2005). Evidence based on interviewing principals (Harris et al., 2010) suggests that principals look for varied sets of personal and professional attributes but tend to try to select teachers that lead to greater diversity on their staffs (in terms of race/ethnicity, gender, and experience).

Some research suggests that hiring processes could be better. An early study by Ballou (1996), for instance, finds that unlike other professions (e.g., law), graduating from a selective college does little to improve one’s job prospects in teaching. By contrast, however, Boyd et al. (2011) find, based on an examination of applications to transfer between schools, that teacher experience is a highly valued attribute. And, Boyd et al. (2013) find that when geographic and school characteristics are accounted for, employers do in fact demonstrate preferences for teachers with stronger academic achievement and higher scores on certification exams. Hinrichs (2013), in a resume study, finds that traditional public schools (but not charter or private schools) have a slight preference for teachers who graduate from more selective colleges.

A few studies look at whether schools identify the best applicants in terms of a direct measure of teacher effectiveness: value-added. Hanushek et al. (2005) analyze whether schools that offer higher levels of compensation (in the form of both salary and work environment) tend to hire more effective teachers. They find that these (generally suburban) schools are more likely to hire teachers with advanced degrees, but find no relationship between hiring and value-added measures of effectiveness. Staiger and Rockoff (2010) also argue that school leaders do not focus on effectiveness during the hiring process. They draw on a natural experiment that occurred in the Los Angeles Unified School District (LAUSD) in 1997, when the state provided a financial incentive to keep K–3 class sizes under 20 children. LAUSD more than doubled its annual hiring of new elementary school teachers during the next five years. During this time, teacher pay did not increase, and the proportion of new hires without teaching credentials increased from 59 to 72 percent. If the district was able to identify the best teachers during the hiring process and had been selecting teachers from the upper end of the applicants’ future performance distribution—that is, selecting those applicants who are likely in the future to be more effective teachers than the nonselected applicants—then one would expect a decrease in the average effectiveness of the cohort of teachers hired in the years after 1997, because the district would need to dip further into the applicant pool to hire enough teachers. Yet, value-added modeling of elementary teacher effectiveness showed no evidence of such a decrease. The results actually suggest that teacher hiring is tantamount to random in terms of future effectiveness, at least as measured by gains in student test scores.

In terms of the hiring process itself, research suggests that new teachers have relatively few interactions with school-based personnel during the hiring process (Liu & Johnson, 2006; Strauss et al., 2000). Few school systems—less than 10 percent in one study (Liu, 2003)—require prospective teachers to demonstrate their craft through the teaching of a sample lesson. Instead, districts rely on transcripts, letters of reference, and resumes (Balter & Duncombe, 2008) and many—perhaps as many as 2,000 school districts, though the actual number is unknown (Metzger & Wu, 2008)—rely on relatively short structured interviews to help make hiring decisions. There is not much evidence suggesting that these teacher interview instruments (e.g., the Urban Teacher Selection “Star” Teacher Interview and Gallup’s Teacher Perceiver Interview protocol) can be used to detect teacher quality.

The lack of information about the effectiveness of the teacher selection processes used by local school systems is troubling. Better knowledge about these processes would provide insights about the merits of tightening or loosening the teacher certification gateway, as well as information that school systems might find useful in making teacher selection decisions. In fact, some evidence suggests hiring could be improved with more information about applicants. Rockoff et al. (2011) examine how traditional information collected by New York City (degree and major, passage of license exam on first attempt, college ranking, SAT scores) and nontraditional information (cognitive ability, Math Knowledge for Teaching, conscientiousness, extraversion, and self-perceptions of general and personal efficacy) corresponds with student and teacher outcomes (including value-added scores, subjective teacher ratings, teacher absences, and teacher retention). They find that few individual teacher attributes are significant predictors of later effectiveness, but that the nontraditional information collectively added considerable explanatory power to predictions: the combination
of both traditional and nontraditional information explains about 12 percent of the expected variance in teacher effectiveness compared to about 4 percent using only traditional information.12

Teacher Workforce Policies

The weak link between paper credentials and student outcomes makes it difficult to affect the teacher quality issue through federal or state policies focused on these credentials. I illustrate this point with an example from my own work on NBPTS certification. As I described above, work with a colleague (Goldhaber & Anthony, 2007) finds that a teacher being NBPTS certified has a statistically significant impact on student achievement and the effect, about 5 percent of a standard deviation, which is nontrivial. But, what does this mean for NBPTS status as a predictor of teacher quality? To answer this question, in follow-up work (Goldhaber, 2006), I estimate the total impact that teachers in the study have on their students and divide the sample of teachers into those who are NBPTS certified and those who are not. This simulation shows that teachers who become NBPTS certified are likely to have larger teacher impacts than the average teacher in the sample, but the effect is not terribly consistent: 59 percent of NBPTS-certified teachers have above-average effects in math, and 56 percent have above-average effects in reading—certainly more than what one would expect from a teacher taken from the workforce at random (50 percent), but far from an overwhelming percentage of the NBPTS-certified population as a whole.

One conclusion that could be drawn from the considerable heterogeneity in teacher effectiveness is that schools ought to craft and emphasize policies and practices that are based on teachers’ in-service practices and performance.13 For example, it is not too controversial to suggest that the type of professional development offered to teachers be tailored to individual teacher needs. More controversial are calls for changes in teacher tenure policies that diminish the up-front (often after three years) teacher job security that exists in many districts (Johnson & Donaldson, 2006). Also contentious is the notion that pay systems ought to be restructured, either so they are more closely aligned to teacher productivity, so-called merit pay (Ballou & Podgursky, 1993; Hanushek & Rivkin, 2003; Podgursky & Springer, 2007; Teaching Commission, 2006), or so that they encourage teachers to acquire key teaching skills, or “knowledge and skills pay” (Odden & Wallace, 2007). But while a number of personnel decisions might be informed by measures of teachers’ in-service performance, it is problematic to act on differences that exist in teacher performance because today’s evaluation systems typically do not differentiate teachers (Weisburg et al., 2009).14 The failure of evaluation systems to recognize the differences that we know exist between teachers (both from statistical analysis and, I would argue, casual observation) is one of the reasons we are now witnessing a major shift in teacher policy in a great many states today, encouraged by federal programs such as the Teacher Incentive Fund and Race to the Top (RTT). See Donaldson and Papay (this volume) for a discussion of recent changes in teacher evaluation. As they note, none of the recent methods—whether they are based on student value added or on classroom observations—are free of error.

Two teacher workforce policies that we know a good deal about are professional development and merit pay. Unfortunately, the empirical evidence about these workforce interventions is not terribly promising. Research on professional development shows that it can, at times and in some contexts, have an impact on teacher performance (Yoon et al., 2007), but, by and large, the typical one-shot workshops that are often offered to teachers have failed to lead to performance changes (Cohen & Hill, 2000; Garet et al., 2001; Kennedy, 1998). Two recent large-scale experimental studies (Garet et al., 2008; Glazerman et al., 2010) offer somewhat conflicting findings about whether professional development leads to statistically identifiable impacts on student achievement.15

Recent evidence on merit pay (in the United States) also suggests it does not generally improve teacher productivity. Specifically, a number of experiments and well-designed nonexperimental studies assess the degree to which monetary incentives for teachers affect their productivity. These experiments have covered several different types of incentives, including those that target individual teachers and those that target teams (typically at the school level). Although in each case the pay system is designed to reward teachers whose students make larger than expected gains on standardized tests, the studies find little detectable impact of
teachers receiving these types of incentives (Fryer et al., 2012; Glazerman & Seifullah, 2010; Imberman & Lovenheim, 2013; Marsh et al., 2011; Springer et al., 2010).

It may be premature to write off professional development and incentives. For instance, as Rice (2009) notes, one reason that professional development might not generally work well is that it is delivered in "... an incentive structure that rewards seat time rather than improved performance. Policymakers need to consider better ways to structure incentives so that teachers engage in professional development directly related to their work with the potential to improve their teaching practices and, ultimately, their effectiveness" (p. 243). Similarly, merit pay might work if it is implemented in the context of a more cohesive set of reforms that includes restructured evaluation and performance feedback mechanisms (Goldhaber & Walch, 2012). Lastly, because some of the research on work-force interventions is designed to find short-term (e.g., over a couple of years) impacts on productivity, it may miss longer run changes in the effectiveness of individual teachers, or changes in the quality of the teacher workforce resulting from a change in the mix of people who enter or exit the teacher workforce (Lazear, 2003). The idea here is that the right policies could increase the likelihood that the more productive teachers stay and less productive teachers leave.

Changing the mix of who is in the teacher workforce is an explicit part of new workforce-oriented policies as a number of new state laws explicitly call for teachers repeatedly deemed ineffective by new evaluation systems to lose either employment protections (tenure status) or their jobs. These reforms are quite controversial and represent a significant departure from teacher policies in the past. Of course it remains to be seen whether new evaluation systems actually lead to these sorts of sanctions for individual teachers given opposition to using evaluations in this way (McNeil, 2013).

A few studies explore the potential effects of changing the mix of people in the teaching profession explicitly through "deselection" policies (e.g., firing, layoffs, nontenuring). The word potential is emphasized because these studies involve the somewhat speculative use of simulations, as opposed to evaluation of policy variation. Yet some of the simulations are quite provocative. Chetty et al. (2011) suggest that replacing a teacher whose true value-added is in the bottom 5 percent with an average teacher would increase students’ lifetime income by $267,000 per classroom taught. Hanushek (2009) concludes that replacing 5 to 10 percent of the least effective teachers (two to three teachers in a school of 30 teachers) with teachers of average effectiveness would put U.S. students in the top tier of nations in terms of international test performance. However, it is hard to determine the downstream effects of workforce policies from simulations, because we do not know how teachers would respond to what would be a radical policy departure from what is currently a very secure occupation. As Rothstein (2013) shows, different assumptions about such behavioral responses imply that workforce policies of this type may in practice have far smaller effects on the quality of the teacher workforce than those predicted by these researchers.

Some early evidence suggests new comprehensive approaches to managing teacher human capital can lead to changes in teacher behavior. A recent study of Washington, DC’s high-profile evaluation and incentive system, known as IMPACT, shows that it affects both teacher behavior and mobility (Dee & Wyckoff, 2013). Specifically, teachers deemed highly effective who were eligible to receive large increases in base pay if they continue to be judged as highly effective were found to increase their performance, and teachers judged to be ineffective, and threatened with dismissal if they were to continue to be found ineffective, were far more likely to voluntarily leave the district.

I believe it makes a good deal of sense, given what research shows about the importance of teacher quality, and the inability of pre-service credentials to accurately predict it, for policymakers to be exploring teacher workforce policy reforms. And while it is too soon to draw any strong conclusions about the efficacy of new comprehensive workforce policies that target in-service teachers, we will no doubt learn a great deal more in the upcoming years as these policies are rolled out.

Notes

1. A one standard deviation increase in teacher quality is estimated to raise student achievement in reading and math by about 10 percent of a standard deviation. See Hanushek and Rivkin (2010) for more detail. Further, the magnitude of the teacher quality effect greatly exceeds other
The Hinrichs (2013) study involves an experiment where schools are sent fictitious resumes with different, randomly chosen attributes, to see if students with NBPTS-certified teachers gain more than other students. For instance, the No Child Left Behind (NCLB) Act specifically denotes that "highly qualified teachers" are those who hold a bachelor's degree, obtain full certification from their state, and demonstrate competency in each subject they teach (as defined by the state). For specific legal definition of "highly qualified teachers," see No Child Left Behind Act, Title IX, Sec. 901, Part A, Sec. 9101.23.

Cavalluzzo, for example, who focuses on the mathematics achievement of high school students in the ninth and 10th grades, finds that, all else equal, students with an NBPTS-certified teacher gain 12 percent of a standard deviation more than other students on a year-end exam. Goldhaber and Anthony study students in grades three to five, and find that having an NBPTS-certified teacher is predicted to improve student achievement by about 5 percent of a standard deviation (in both subjects). They estimate that having a teacher at the 85th percentile of the teacher performance distribution rather than at the mean is worth an additional $130,000 in combined discounted lifetime earnings to the students in the classroom.

Specific value-added models may differ in their details, but the basic idea behind them is to translate measures of student achievement gains on standardized tests into a measure of teacher performance. For more background on how specific models function to translate measures of student achievement into measures of teacher effectiveness (and the implications for teachers instructing different student populations), see Goldhaber et al. (2013).

For instance, the No Child Left Behind (NCLB) Act specifically denotes that "highly qualified teachers" are those who hold a bachelor's degree, obtain full certification from their state, and demonstrate competency in each subject they teach (as defined by the state). For specific legal definition of "highly qualified teachers," see No Child Left Behind Act, Title IX, Sec. 901, Part A, Sec. 9101.23.

See, for instance, Aaronson et al. (2007), Clotfelter et al. (2010), Goldhaber and Hansen (2013), Kane et al. (2008), Harris and Sass (2011), Rivkin et al. (2005), and Rockoff et al. (2011).

These findings have also been found using nonexperimental research methods (e.g., Kane et al., 2008; Xu et al., 2011). Not all studies, however, reach this conclusion. Boyd et al. (2009) find that TFA teachers are slightly less effective than teachers with traditional licensure in both math and reading. For recent reviews on this issue, see Feuer et al. (2013) and Goldhaber (2013). Some recent research (e.g., Boyd et al., 2009) has attempted to address this knowledge gap by focusing on whether any teacher training program characteristics (e.g., timing of student teaching) predict effectiveness, and/or whether student experiences (Ronfeldt, 2012) affect it.

Not all of these other teachers were fully licensed, but this is arguably the right comparison group given that they are the teachers hired into those schools also employing TFA teachers. Comparable findings have also been found using nonexperimental research methods (e.g., Kane et al., 2008; Xu et al., 2011). Not all studies, however, reach this conclusion. Boyd et al. (2009) find that TFA teachers are slightly less effective than teachers with traditional licensure in both math and reading. For recent studies that compare the outcomes of teachers graduating from different training institutions, see Boyd et al. (2009), Goldhaber et al. (2013), Koedel et al. (2012), Mihaly et al. (2012), and Ronfeldt (2012).

For recent reviews on this issue, see Feuer et al. (2013) and Goldhaber (2013). Some recent research (e.g., Boyd et al., 2009) has attempted to address this knowledge gap by focusing on whether any teacher training program characteristics (e.g., timing of student teaching) predict effectiveness, and/or whether student experiences (Ronfeldt, 2012) affect it.

To put the experience and other effects into context, as students move from one grade to the next, they typically gain about 1 standard deviation in math and reading achievement in the lower elementary grades, one-half a standard deviation in the upper elementary grades, one-third of a standard deviation in middle school, and a one-quarter of a standard deviation in high school (Bloom et al., 2008); and the average gap between black and white students, or economically advantaged and disadvantaged students, is on the order of magnitude of 0.7 to 1.0 standard deviations (Hanushek & Rivkin, 2010).

There is less empirical evidence examining the relationship between subject-matter training of teachers and student achievement at the elementary level, and the existing evidence is less conclusive. Eberts and Stone (1984) address this relationship and find no statistically significant relationship between the number of math courses taken by teachers and their fourth-grade students' achievement in math. Similarly, Hill, Rowan, and Ball (2005) do not find a relationship between the number of math methods and content courses (they group these together) a teacher has completed and student gains in mathematics (in the first and third grades).

This finding is not universal. Several recent studies, for instance, fail to find evidence supporting a relationship between subject-specific training and student achievement at the high school level (Aaronson et al., 2007; Harris & Sass, 2011). A possible explanation for the divergent findings is that the quality of the master's degrees that teachers are receiving has declined in recent times with the huge increase in the last decade in the proportion of master's degrees that are being granted by online for-profit colleges (Sawchuk, 2013). While there is little empirical evidence about these institutions (Lang & Weinstein, 2013), their quality has been questioned (U.S. Senate Committee, 2012).

Cavalluzzo, for example, who focuses on the mathematics achievement of high school students in the ninth and 10th grades, finds that, all else equal, students with an NBPTS-certified teacher gain 12 percent of a standard deviation more than other students on a year-end exam. Goldhaber and Anthony study students in grades three to five, and find that having an NBPTS-certified teacher is predicted to improve student achievement (outside of the year that a teacher becomes certified) by about 5 percent of a standard deviation (in both subjects).

It would not be surprising, for instance, to find cohort effects given that NBPTS certification was relatively new in the mid- to late 1990s, but it is also the case that the NBPTS assessment process has changed over time (e.g., the weighting of different assessments used to judge NBPTS candidates changed in 2001).

The Hinrichs (2013) study involves an experiment where schools are sent fictitious resumes with different, randomly chosen attributes, to see...
which applicant attributes appear to be desirable.

17. Note also that Dobbie (2011) finds a statistically significant relationship between subjective assessments of TFA applicants and the subsequent achievement of students with TFA teachers in their first year of teaching.

18. There is evidence that measures of teachers’ in-service performance, value-added measures in particular, predict later student achievement better than the credentials that are typically used for high-stakes personnel decisions (Goldhaber & Hansen, 2010). For a discussion of reliability and validity issues associated with using value added to assess individual teachers, see Rothstein (2010) and Kane et al. (2013).

19. This finding was documented in The Widget Effect, which studied the evaluation practices in 12 school districts across four states. The study found the frequency and methods of teacher evaluation varied, but the results of evaluations were almost always the same: Over 99 percent of teachers in districts with a binary evaluation system are rated as satisfactory, and over 94 percent of teachers in districts with a multitiered rating system receive one of the top two ratings, with less than 1 percent of teachers receiving an unsatisfactory rating.

20. A randomized control trial focusing on a 1-year content-focused professional development (PD) program thought to be well designed showed positive impacts on teachers’ knowledge of scientifically based reading instruction and instructional practices promoted by the PD program, but no effects on student test scores (Garet et al., 2008). Another recent large-scale experiment did find evidence that teachers who received two years of comprehensive induction had higher levels of student achievement in the third year, but the results were puzzling in that there were no discernible effects from the first two years of receiving this support (Glazerman et al., 2010).

21. One notable exception to this is Fryer et al. (2012), who investigate two kinds of incentives: the traditional type whereby teachers who have students meeting achievement objectives receive performance incentives, and a second type whereby teachers are provided with incentives in advance and will lose the incentive if their students do not meet specified achievement objectives (testing “loss aversion”). He finds no statistically significant results from the traditional type of incentive, but large significant effects associated with the threat of losing bonuses received in advance. While this study suggests the potential to change the productivity of in-service teachers, it is difficult to imagine how this type of incentive could be used in public schools at scale.

22. Behavioral responses to performance information, and the use of evaluations, may be influenced by the nature of the system. While the validity and reliability of performance evaluations are no doubt important, issues of trust, understanding, and emotional salience are also likely to influence the way teachers react to the feedback they receive about performance.

23. The authors are careful to note that these findings (which are based on a regression discontinuity analyses) should not be interpreted as average treatment effects since they apply only to teachers who are near to the high-stakes cutoff points.

References


Recognizing that, of all school factors, teachers contribute the most to student achievement, U.S. education policy in the past decade has focused on improving teacher quality (Goldhaber, this volume). Policymakers have seized on teacher evaluation as a primary lever for achieving this goal. Since 2008, economic and political factors have converged to elevate teacher evaluation on the federal, state, and local education agendas (Donaldson & Papay, 2014; McGuinn, 2012). Through Race to the Top and No Child Left Behind waiver requirements, the federal government has pushed states to reform their teacher evaluation systems. Between 2009 and 2012, the majority of states and Washington, DC, responded by passing new legislation to develop new statewide evaluation systems or specify requirements for districts to follow in evaluation reform (National Council on Teacher Quality, 2013). This chapter presents what is known about current teacher evaluation initiatives.

Reforms to teacher evaluation have a long history, and current debates reflect longstanding disputes. While today’s interest in teacher evaluation echoes prior waves of investment in this policy lever, contemporary reforms are more widespread, affecting almost all teachers in the United States, and are generally more systematic and centralized. Most of the new evaluation systems increase the frequency of evaluations and include more summative rating categories than did their forerunners. Furthermore, many new systems rely on mechanisms that have not been widely used to evaluate teacher effectiveness in the past, namely student achievement data and standards-based observations of teacher practice.

Like all performance appraisal systems, teacher evaluation can serve different purposes. We focus on two purposes that inform most evaluation systems: accountability and development. First, an evaluation system can provide accountability through monitoring, ensuring that teachers exert sufficient effort and follow specific standards and practices. It can also provide accountability by serving as a basis for rewarding or sanctioning teachers based on their performance. With better information about teacher performance, administrators can identify persistently low-performing teachers and dismiss them or counsel them to resign, and they can provide monetary or other rewards to high-performing teachers. Second, a strong teacher evaluation system can support teacher instructional development, offering actionable feedback to teachers about their practice and supporting them in improving their skills. Most evaluation systems embody these dual purposes.

While these development and accountability goals can complement each other, they can also produce tensions. In practice, evaluation systems emphasize either accountability or development, whether explicitly or implicitly, with substantial implications for the system’s design, the measures it incorporates, and the infrastructure that undergirds it. For example, 42 states report in their NCLB waiver applications or their state teacher evaluation guidance that the goal of evaluation reform in their state is professional learning (Center on Great Teachers and Leaders, 2014). To date, though, states have struggled to create systems that enable teachers to translate feedback into lasting changes in their instruction. Moreover, policy and research have focused on ensuring the reliability and accuracy of ratings of teacher performance for accountability, with little attention to how such measures can promote development (Papay, 2012; Donaldson et al., 2014; Firestone et al., 2013; Sartain et al., 2011).

We focus attention on three widely used measures in evaluation reform today: standards-based observations, value-added measures, and student learning objectives. Standards-based observations use data from classroom observations to assess how well teachers are performing on specific instructional standards adopted by the district. Value-added measures use statistical analysis of students’ test-score histories to estimate a teacher’s contribution to student achievement. Student learning objectives also use test scores, but evaluate a teacher’s
progress relative to goals that she and her evaluator have set together. Each measure has advantages and disadvantages for serving the dual purposes of teacher evaluation.

A Brief History of Teacher Evaluation

Since 1900, teacher evaluation policy has transitioned through several main phases. Efforts to urbanize public schools in the progressive era of the early 20th century launched the first wave of teacher evaluation reform. At this time, teaching was largely viewed as a short-term career without a strong professional core of knowledge (Lortie, 1975), and formal evaluation of practice was essentially nonexistent. Building on the scientific management ideas promoted by Frederick Winslow Taylor and his adherents (Taylor, 1911), school reformers argued for increased monitoring of teachers’ instruction by recording the type and duration of teachers’ specific activities on rating scales (Hanus, 1914). For example, an early 20th century “teacher efficiency score card” had supervisors rate teachers on attributes such as “general appearance” and “voice” (Cubberley, 1929). Evaluators were to use these ratings to evaluate a teacher’s efficiency, thereby revealing areas of weakness and providing teachers with motivation to improve (Elliott, 1915). These efforts dovetailed with those of early educational psychologists, such as Thorndike (1913), who focused on the psychometric properties of such ratings.

These rating systems provoked arguments similar to today’s debates. In these systems, very few teachers received low ratings (Hanus, 1914) and leaders were concerned about whether principals had sufficient time to supervise well. Moreover, debates raged about how to use the results of evaluations, including whether to pay teachers for their performance. Critics during this period argued that effective teaching, unlike factory work, did not lend itself to such concrete measures for teacher accountability. Hunter (1924) opposed the “inspection” orientation of the period, and advocated instead for efforts to promote professional growth.

After World War II, the literature on teacher evaluation reflects a shift from inspection for accountability towards supervision for teacher development. In the 1950s, evaluation focused on building relationships between teachers and leaders, and in the 1960s, scholars and practitioners sought to codify these efforts into the practice of clinical supervision. This new work in evaluation continued to sidestep the calls to define effective practice and rate teachers accordingly that had been popular earlier in the century. Instead, clinical supervision specified the process for evaluation. Goldhammer and colleagues’ (1969) five-stage process, including observations and pre- and post-observation conferences, is still the basis for many evaluation processes today. Thus, with minor exception in some innovative districts, by the 1970s teacher evaluation across the country tended to follow a fairly standard pattern.

The 1980s and 1990s witnessed growing interest in teacher evaluation as a lever for school improvement (Darling-Hammond et al., 1983), in part due to calls to pay teachers for their performance or promote them on career ladders based on their evaluation results (Murnane & Cohen, 1986). Beginning with Hunter’s Mastery Teaching (1982), practitioners and researchers began to specify, in some detail, the types of practices that constitute effective teaching. Building on the work of Hunter and others in the 1980s, several efforts throughout the 1990s helped to define rigorous, detailed standards for effective teaching practice, with corresponding evaluation rubrics. Charlotte Danielson’s Framework for Teaching, the National Board for Professional Teaching Standards, and others grew in popularity as a basis for revised evaluation systems.

Nonetheless, by the early 2000s, teacher evaluation in most settings remained sporadic, with little influence on either teacher accountability or development. However, some districts had used the new ideas to make substantial changes to evaluation practices. Districts such as Cincinnati, OH, and Montgomery County, MD, had developed new evaluation systems through collective bargaining in the 1990s. Both systems adopted new standards of teaching practice that formed the basis for professional evaluation. Cincinnati’s approach was adapted from Danielson’s Framework and Montgomery County’s derived from the National Board. While they provided support and training for raters—including peer teachers—to ensure scoring accuracy, both systems incorporated explicit strategies for professional growth. Both districts also created Peer Assistance and Review programs in which expert teachers provided intensive evaluation and support for novice teachers and struggling veterans, coupling a serious accountability process that could lead to teacher dismissals with a robust focus on professional learning. These districts—and others—became models of reform for other states.
and districts (Johnson et al., 2010; Johnson, 1988; Koppich, 2012; Johnson et al., 2009).

The first decade of the 2000s saw growing acknowledgment that these pockets of reform were not sufficient and that the traditional evaluation system present in many districts was inadequate. As Randi Weingarten, president of the American Federation of Teachers, put it, “Our system of evaluating teachers has never been adequate” (2010, p. 3). Echoing calls in earlier eras for evaluation reform, reports such as Toch and Rothman’s Rush to Judgment (2008) and the New Teacher Project’s Widget Effect (Weisberg et al., 2009) brought into stark relief the limitations of these approaches. These reports, and other studies, documented that evaluations in such systems were rare, brief, and provided little information to teachers or districts regarding teachers’ instruction. The lack of effective evaluations led to a “Lake Wobegon” problem where nearly all teachers were deemed “satisfactory” and, despite the recognition that some teachers might not be successful, almost no teachers were dismissed, even when policies removed barriers to dismissal (Donaldson, 2009; Jacob, 2011; Weisberg et al., 2009; Tucker, 1997; Honawar, 2007). In essence, teacher evaluation had failed on a broad scale to accomplish its accountability purpose.

Furthermore, evaluation was largely ineffective in promoting teacher development. Despite the importance of feedback to teachers and its potential to improve their performance, the traditional evaluation process often provided little of it (Weisberg et al., 2009). When feedback was provided, it often lacked specifics (Sinnema & Robinson, 2007), failed to promote deep reflection or analysis by teachers (Sartain et al., 2011), and was only loosely connected to professional learning (Stronge & Tucker, 2003).

While these types of limited evaluation systems—where teachers are not evaluated regularly, and the evaluations that do occur are cursory at best—still exist in some districts across the country, the past five years have ushered in dramatic attempts to change the status quo. Three key stimuli have helped promote this new wave of evaluation reform. First, as discussed above, researchers and practitioners have worked to develop clearer standards of effective instructional practice. Beyond simply specifying standards, these frameworks provide detailed descriptions of the practices exhibited by teachers at different levels of professional mastery. These explicit descriptions enable evaluators to distinguish among teachers for both accountability and development purposes.

Second, the advance of standards-based reform and annual student assessments nationwide raised the stature of teacher evaluation in several ways. Increasingly comprehensive databases on student performance supported a growing number of influential studies in the early 2000s documenting that teachers are the most important school-level factor in promoting student achievement and that teachers vary widely in their abilities to raise student test scores (see, e.g., Aaronson, Barrow, & Sander, 2007; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). These studies spurred growing interest in using student test scores to evaluate teacher practice, as researchers worked to develop increasingly sophisticated statistical models that attempted to isolate an individual teacher’s contribution to students’ scores on state tests.

Third, the growing role of the federal government in education has led to more federal influence over teacher policy. No Child Left Behind set a key precedent for increasing the federal government’s role in holding schools accountable for student performance. Under President Obama, the federal government took an even larger role, and the focus has shifted even more from the school level to the teacher level. Recent policies such as Race to the Top, No Child Left Behind waivers, and Teacher Incentive Fund competitions have spurred the creation of new state and district policies and programs focused on teacher evaluation (McGuinn, 2012). Simultaneously, political and economic developments made states and districts more tolerant of an expanded federal role and raised the prominence of teacher evaluation (Donaldson & Papay, 2014; McGuinn, 2012). Threatened by eroding public support and well-orchestrated public critiques, teachers unions have become more willing to collaborate with districts on matters of teacher quality, including reforming evaluation systems. Weingarten (2010), for example, has repeatedly advocated for more thoughtful and rigorous teacher evaluation in recent years. At the local level, an increasing number of union leaders have made clear that, while they must uphold due process requirements, they will not blindly support any teacher who is dismissed. These union presidents are also, by and large, willing to work with district leaders to pursue reforms in evaluation and other teacher policies, in addition to advocating for better salaries and benefits (Johnson et al., 2009).
Teacher Evaluation Frameworks and Components

As a result of these trends, many states and districts have begun to develop new teacher evaluation models over the past five years. Like their predecessors, these systems are built on the premise that teacher evaluation can improve instructional quality and enhance student learning by holding teachers accountable and providing development opportunities. However, systems differ in their relative emphasis on these two purposes.

As states and districts reform teacher evaluation practices, they face several important decisions. Central among these is their choice about what measures, or combination of measures, to use. These decisions require policymakers to attend to the psychometric properties of the measures—including the extent to which they are unbiased and precise/reliable—as well as their validity for evaluating teachers and the types of feedback they provide. This choice should align with the primary purpose of evaluation in the state or district. In this section, we discuss the strengths and weaknesses of these measures, particularly as they relate to accountability and teacher development.

Standards-Based Classroom Observations

As of 2013, 44 states required that classroom observations be part of teacher evaluation ratings, making it the most commonly used component of teacher evaluation systems (NCTQ, 2013). Unlike in the past, observation today typically involves assessing a teacher’s instructional practices against an articulated set of performance standards. Formal observations are generally structured around an observation instrument or protocol that helps evaluators measure teachers’ performance by translating observed practice on specific standards into detailed descriptors within a rubric. Evaluators may observe for an entire class period or “walk through” a classroom to observe only briefly. Informal “walk-through” observations are typically less structured than are formal ones.

One example comes from Cincinnati, which implemented standards-based observations in 2000 based on a modified version of Danielson’s Framework for Teaching. The district has 32 performance elements organized into 16 instructional standards. Trained evaluators observe individual teachers several times during the year, scripting lessons and matching evidence of classroom practice to performance levels on the rubric for each standard. In addition to detailed information about performance on each of these standards, evaluators give each teacher a summary rating of “distinguished,” “proficient,” “basic,” or “unsatisfactory” (Kane et al., 2011). The instructional standards explicitly examine teachers’ instructional interactions with students and students’ responses to their teachers (Johnson et al., 2009).

Like Cincinnati, most states and districts that have developed professional standards for teachers typically adapt their standards and protocols from existing models. The Danielson Framework is the most popular model, but several other examples of standards and observational rubrics exist. Of these, the National Board for Professional Teaching Standards (NBPTS) are also used regularly, while some states and districts have sought to incorporate work from more specialized standards and observational instruments for English language arts (Protocol for Language Arts Teaching Observation, PLATO), mathematics (Mathematics Quality of Instruction, MQI), and classroom interactions (the Classroom Assessment Scoring System, CLASS). Some observers argue that observational rubrics need to be subject specific, as the competencies and practices necessary to succeed with students vary by subject area (Hill & Grossman, 2013).

To be effective evaluation tools, these standards-based classroom observations must not only differentiate the quality of instructional practice, but also explore how teachers use these practices to promote student learning. In other words, an effective standards-based observation system will evaluate not only a teacher’s actions but also teacher-student interactions. While standards-based observations may measure student learning as viewed during the observed class period and some observational protocols include consideration of student work, this approach focuses more directly on teacher inputs than student outputs.

Existing research on these standards-based protocols documents widespread variation in teacher practice and shows that teachers who score more highly on these rubrics also have greater estimated contributions to student test scores (i.e., value-added). Although test scores and observations may measure different dimensions of practice, examining their relationship proves useful. Indeed, standards-based observation ratings actually
predict student test performance above and beyond that teacher’s value-added rating (Kane et al., 2011; Kane et al., 2013).

In theory, standards-based observations work quite well as accountability mechanisms for monitoring, as trained observers can examine directly the degree of teacher compliance with district expectations for instruction as defined in the observation rubric. The real constraint in this regard is cost, as observations of full class periods take substantial human resources to complete. Here, briefer walk-throughs might be warranted, particularly if coupled with more substantial observations at several points during the year. We return to the issue of cost below.

As tools for developing a clear measure of teacher productivity for accountability sanctions and incentives, standards-based evaluations raise several concerns. First, they rely on human judgments. Evaluators report that it is difficult to separate what they know of the teacher, or the teacher’s contributions outside of the classroom, from their judgments of the teacher’s instructional practice (Donaldson, 2013; Papay & Johnson, 2012). Administrators consistently assign higher evaluation ratings to teachers than do external observers. Moreover, they also rate teachers higher on their evaluations than on informal effectiveness ratings provided to researchers (Sartain et al., 2011; Harris et al., 2014). Although having clear standards, using highly qualified and well-trained evaluators, and focusing on evidence can help remove much of the subjective bias, separating the personal from the professional can be difficult. However, standards-based observations and subjective evaluations allow supervisors to take into account a much wider range of data, thereby making potentially stronger inferences about a teacher’s true contributions to student learning than can measures based on test scores.

A second consideration for standards-based observations as accountability measures is that they require evaluators to make judgments based on a relatively limited sample of instruction, which can constrain the reliability of these measures. Also, because different evaluators may have different standards, achieving and maintaining sufficient inter-rater reliability may be difficult. Hill and colleagues (2012) argue that reliability is not simply a product of the instrument, but also of context, rater training, and scoring designs. Evaluators are typically school principals, who have many other responsibilities and often struggle to find sufficient time to do this work well. Although building and maintaining reliability takes a substantial investment, it is possible. Incorporating multiple observations into an evaluation helps a great deal, particularly if the observations are unannounced. For example, in Cincinnati, evaluators generally observe at least four complete lessons before making their final determinations, and all evaluators must complete a rigorous training program, earn certification as reliable observers, and participate in ongoing training and professional development for evaluators (Johnson et al., 2010; Kane et al., 2011). Research also suggests that using multiple observers can resolve some of these potential reliability challenges (Hill et al., 2012; Kane et al., 2013). To resolve the cost implications of this decision, some districts (for instance New Haven, Connecticut) require multiple raters only for teachers with very high or very low ratings that could result in rewards or sanctions (Donaldson & Papay, 2014).

For the purpose of teacher development, standards-based observations have strong potential. They can provide direct and specific feedback about a teacher’s instructional practice. The observation reports and summative evaluations include detailed information about teachers’ strengths and weaknesses and thus can be meaningful to teachers and administrators. The performance standards lay out overall expectations for performance and rubrics and describe a continuum of instructional quality, thereby offering teachers a clearer sense of the practices they need to adopt in order to improve their teaching. Outside of the evaluation context, the rubrics and standards can also provide fodder for broader conversations between administrators and teachers and among teachers about how to improve instruction (Donaldson et al., 2014). In this way, evaluation systems with standards-based observations potentially enhance teachers’ skills, thereby serving evaluation’s development purpose.

**Value-Added Measures**

Over the past decade, the federal government has pushed states and districts to include student achievement measures in teacher evaluation, and states have responded: 41 states require that teachers be evaluated at least in part on student achievement (NCTQ, 2013). Value-added measures, which use test scores to estimate a
teacher’s contribution to student achievement, are a popular but controversial option. Many researchers, primarily economists, see them as useful components of a teacher evaluation system (Hanushek & Rivkin, 2010; Glazerman et al., 2010; Kane & Staiger, 2012), largely because they turn the conversation from educational inputs to educational outputs. In theory, supervisors want information about teacher productivity, and value-added methods can provide some evidence in this regard. Others view them as a cause for concern (Baker et al., 2010), primarily because they may not account well enough for differences in the types of students that teachers teach and may not be sufficiently reliable to estimate accurately a teacher’s contributions to student learning. In large part, this debate depends on how value-added data will be used to hold teachers accountable.

Many types of value-added models exist: the most well-known is the Education Value-Added Assessment System (EVAAS) developed by William Sanders, currently in use in districts in at least 20 states. The central idea behind value-added approaches involves comparing the test scores of a teacher’s students to the predicted scores had the students had an “average” teacher. In this regard, models that use student growth percentiles are quite similar to value-added models as we discuss them here. Analysts make a variety of decisions in estimating these predicted levels of achievement, such as whether to account for student characteristics and whether to compare teachers in the same school or across the district. These are, in effect, policy decisions without clear answers. For example, if the models account for the disadvantages of student poverty, they essentially hold the teachers of low-income students to different standards than those of higher-income students. However, if they do not account for poverty, they may not fully control for out-of-school factors affecting students.

In the past decade, substantial research has examined value-added models. This research asks two basic questions: (1) on average, are teachers with higher value-added estimates more effective? and (2) for an individual teacher, do value-added models provide an unbiased and reliable estimate of their effectiveness? By and large, research finds that, on average, students whose teachers have higher value-added scores have better long-term outcomes in the form of a greater probability of going on to college, attending colleges of higher quality, and earning more as adults (Chetty, Friedman, & Rock-off, 2013; Chetty et al., 2011). This work suggests that teachers indeed contribute to outcomes, such as lifetime earnings, that are ultimately far more vital than student test scores. Importantly, though, they capture only one dimension of a teacher’s performance: Jackson (2012) finds that measures based on test scores alone understate the total effect of teachers on students.

While these findings suggest that value-added measures, on average, can identify meaningful performance differences among teachers, estimates for individual teachers can vary widely, for example, across the specific statistical model or test used, or over time. Two central issues emerge here. First, do value-added models provide an unbiased causal estimate of a teacher’s contribution to student achievement? Second, are value-added estimates reliable enough to provide sufficiently precise assessments of teacher effectiveness?

In terms of bias, value-added models are based on standardized tests, which are typically graded by machine or by external scorers with no chance for systematic favoritism towards certain teachers. But simply removing such favoritism does not mean that value-added models are free of bias. Researchers have examined whether value-added models can fully account for differences in student assignments across individual teachers, but this debate has not been resolved. If value-added measures are to isolate a teacher’s contribution to student achievement growth, they must fully account for any differences in the types of students taught, both within and across schools. The MET Study (Kane et al., 2013) and Chetty et al. (2013) conclude that the results from the value-added measures approximate unbiased estimates, while Rothstein (2010) finds that value-added models cannot fully account for the tracking of students within schools. In practice, these disparate results may arise from differences in the patterns of student-teacher sorting across school districts and in the information available in datasets to incorporate in the models themselves.

However, regardless of whether value-added models can fully account for the nonrandom assignment of students to teachers, a broader (and less studied) question is whether value-added models can isolate an individual teacher’s causal impact on student performance (Reardon & Raudenbush, 2009). For instance, attributing a student’s mathematics test gains to her mathematics teacher may not fully recognize the contribution that a science teacher made to the student’s mathematical knowledge. And, it is often difficult to determine which students should be attributed to which teacher, because of within-year student mobility, formal or informal team teaching, and data system limitations.
In practice, a larger concern may be the precision of these estimates. Value-added estimates necessarily include some margin of error around a teacher’s score, and in many cases, this error is sufficiently large that only the very top and bottom performers can be differentiated statistically from the average teacher. As a result, researchers, even those who advocate the use of value-added approaches, often call for using multiple years’ worth of data to construct an estimate for an individual teacher (Glazerman et al., 2010; Braun et al., 2010; Baker et al., 2010).

These issues all contribute to some potentially problematic results for using value-added estimates for teacher accountability. Researchers have examined value-added estimates for the same teacher using different tests within one subject and for the same teacher with the same test from year to year. Although the results differ by dataset and model used, most studies find moderate-sized correlations across different tests or years for the same teacher that average between 0.3 and 0.5 (Corcoran, Jennings, & Beveridge, 2011; Papay, 2011; Goldhaber & Hansen, 2008), suggesting that many teachers would be classified differently based on these estimates.

As a measure to inform teacher development, value-added measures also have substantial limitations. Teachers typically receive a summative score, with no information about their performance on specific areas of instructional practice and no targeted feedback about how to improve. Thus, value-added, as currently conceived, is not well-aligned with evaluation as a tool for professional growth.

In addition, several practical limitations also restrict the broad-based application of value-added models regardless of the evaluation purpose. Most obviously, given standardized testing practices in most states, only a minority of teachers work in a grade or subject area that supports value-added analysis (typically English language arts and mathematics teachers in grades four to eight). In Washington, DC, for example, only 17 percent of general education teachers had value-added scores (Dee & Wyckoff, 2013). Second, these estimates are only as good as the tests on which they are based. Because many state tests are designed to measure whether students are proficient on a set of standards, they may not be particularly good for evaluating teachers. Student performance on tests likely does not reflect the full range of learning that educators and parents care about. Finally, because value-added estimates rely on the test data, principals need to wait at least until test results are available during the summer before assessing performance. And any efforts to include multiple years of data to have sufficiently reliable estimates further limit their use for annual teacher evaluation.

**Student Learning Objectives (SLOs)**

Some districts and states seeking an alternative to value-added for incorporating student achievement data in teacher evaluation have adopted student learning objectives (SLOs). SLOs are goals set by teachers and administrators for each classroom based on existing assessments. Teachers are then evaluated based on how well their students meet these goals. The key, of course, is in how goals are set. Importantly, the rigor of a teacher’s goals depends both on the individual teacher and on her interactions with the principal, not some external standard.

The use of SLOs to evaluate teachers is relatively new. Denver, Colorado, became one of the first districts to use SLOs in 1999 when it adopted them as a key component of ProComp, its performance-based pay system for teachers. SLOs have been adopted as evaluation measures in Austin, Texas; New Haven, Connecticut; Indiana; Maryland; New Jersey; Nebraska; Rhode Island; and Connecticut (Hull, 2013). However, little research has explored their use in teacher evaluation (Goe & Holdheide, 2011).

Although the use of SLOs within teacher evaluation systems is increasing, they do not necessarily perform well as tools for holding teachers accountable. Studies of Denver’s four-year pilot program using SLOs provide some of the only evidence on the measurement properties of this type of instrument. Goldhaber and Walch (2012) found that teachers who received a bonus for meeting their SLOs under ProComp were slightly more likely to fall in the highest two quintiles of the value-added distribution of teachers than in the bottom two quintiles. In mathematics, 42 percent of teachers who received the SLO bonus were in the top two quintiles of the distribution of effectiveness, compared to 37 percent in the lowest two quintiles. In reading, the figures were 41 and 38 percent, respectively. Slotnik and Smith (2004) found that students’ mean growth scores on standardized tests were higher in classrooms of teachers who met both their SLOs as compared to teachers who
did not. However, the study also found several problematic results. For example, over time, teachers became better able to meet the goals they had set, with 98 percent of experienced teachers meeting both their objectives during the final year.

Nonetheless, SLOs have several strengths. First, unlike value-added measures, SLOs can apply to all teachers and all specialists. Second, goal-setting theory suggests that individuals are motivated by goals that they set for themselves (Locke & Latham, 2002). This finding is most relevant to the use of SLOs as development tools, because they may encourage teachers to build skills and improve their instruction. Teachers in New Haven reported that SLOs were the most positive aspect of their new evaluation system and helped them focus their instruction on student performance (Donaldson, 2013). Furthermore, teachers report that the adoption of a new evaluation system incorporating SLOs led them to spend more time analyzing student data and assessing student progress than they had prior to the new system (Donaldson et al., 2014). Third, these measures, which are jointly constructed by teachers and administrators, may increase teachers’ buy-in and investment in their own evaluation. Briggs (2013) argues that because teacher buy-in is necessary for the success of teacher evaluation systems, teachers should have control over some of the ways in which student test data are used to assess them.

Other Means of Evaluating Teachers

Although standards-based evaluations and test-score measures have received the most attention from researchers and policymakers, other means of evaluating teachers are gaining some currency in systems across the country. Research on these measures and their potential to support teacher accountability or development is limited. Student surveys are being used in an increasing number of districts (Hull, 2013). Research has found that students are able to distinguish a teacher whom they like from one who promotes their learning and that student perceptions are correlated with teacher performance (Peterson et al., 2001). In the MET study (2010), student perceptions of instructional quality predicted large differences in student achievement gains and, the authors point out, provided teachers with feedback on how to improve their practice. Some districts are also administering parent/caregiver surveys. Parental involvement in teacher evaluation dates back to the colonial period, although today’s efforts are more systematic and standardized than in the past. States such as New York, Connecticut, and Utah and districts such as Miami-Dade, Florida, mandate the use of parent surveys/feedback in teacher evaluation. The little research that exists suggests that parent survey data may contribute to improving measures of teacher performance (Peterson et al., 2003).

Some districts include portfolios, consisting of teachers’ instructional materials and student work, in teacher evaluation. U.S. Secretary of Education Arne Duncan has publicly recognized Tennessee’s Fine Arts Growth Measures System as a model for how states can evaluate teachers of nontested subjects such as art (Robelen, 2013). Much of the research on portfolios has investigated the relationship between National Board certification, which uses a portfolio-based process, and student achievement. The evidence is mixed, as the effectiveness of Board-certified teachers seems to vary by state (Goldhaber & Anthony, 2007; Clotfelter, Ladd, & Vigdor, 2007; Harris & Sass, 2009).

Combining Measures

In many districts and states that are adopting new evaluation systems, a teacher’s final rating derives from some combination of different measures of teacher effectiveness. Good evidence suggests that using multiple measures can provide a better assessment than using student test score measures alone (see Kane et al., 2013). However, research does not provide strong guidance about how best to combine these measures for different purposes of evaluation (Glazerman et al., 2010; Special Issue: Multiple Perspectives on Multiple Measures, 2003). As such, districts that have adopted multiple measure systems have done so in a relatively ad hoc way. For example, the IMPACT system in Washington, DC, gives each teacher a rating based on 35 percent for value-added, 40 percent for standards-based observation, 15 percent for student learning objectives, and 10 percent for demonstrated commitment to the school community. This distribution of weights represents a marked change from the program’s earlier implementation, when value-added accounted for 50 percent of the teacher’s
score and student learning objectives were not included. For the 80 percent of teachers who work in nontested grades and subjects, standards-based observations represent 75 percent of the total rating.

It is likely that several key influences have encouraged policymakers to adopt such practices. First, there is some general understanding of the value of multiple measures. Second, the federal government’s push to have states include student achievement measures in evaluation, coupled with local political pressures that teacher evaluation not be based solely on test scores, led states and districts to develop systems that use a combination of measures. Whatever the reason, the extent to which policymakers have carefully considered the purposes of evaluation and aligned their use of multiple measures to their preferred purposes is less clear.

Implementation Considerations

Generally, policymakers and researchers have focused on selecting measures, deciding how best to combine them, and determining ways to ensure accurate ratings based on them. However, the implementation of new evaluations composed of the above components also necessitates a number of practical considerations, including what the system will cost and who will serve as evaluators.

Costs

Although almost no research exists on the cost of new evaluation systems, implementing a more rigorous system based on the above components cannot be done cheaply. Planning and development take time and money, and substantial work must be undertaken to ensure that accountability and development purposes are achieved and teachers receive accurate ratings, understand their ratings, and have good information about how to change their practices to improve their rating. However, different elements introduce different types of costs to the school system.

Clearly, operating a rigorous standards-based evaluation system is quite resource intensive. Many of the flaws of prior systems arose because teachers were rarely evaluated. Simply adopting rigorous standards and protocols will not change this practice. To carry out new evaluation programs, school systems must develop substantial capacity. Here, evaluators need time and training to carry out evaluation effectively. In districts that use administrators as evaluators, that means providing additional support to relieve them of some daily administrative tasks.

In one of the few examinations of the costs of teacher evaluation, Papay and Johnson (2012) found that the seven districts in their sample spent between $3,000 and $10,000 per teacher to fund their Peer Assistance and Review (PAR) evaluation programs. It is important to note that the peer assistance aspect makes these programs more intensive than typical teacher evaluation systems. Analyzing the costs and benefits of PAR in these districts, the authors conclude that the program may in fact save districts money in the long run by reducing dismissal costs and increasing the retention of promising new teachers (2012, p. 723). Taylor and Tyler (2012) also discuss briefly the costs associated with the Cincinnati Public Schools’ teacher evaluation system, which uses standards-based observations and peer assistance and review. They report that between 2004–05 and 2009–10, the district allocated between $1.8 and $2.1 million annually ($7,500 per teacher) to pay the salaries of evaluators, including teachers with full-time evaluation duties, and other direct costs. Both of these studies note that the opportunity costs of evaluator’s time may be much larger than the direct costs, which is particularly relevant if expert teachers are pulled out of the classroom to conduct evaluations.

Importantly, while personnel costs are much smaller for value-added models or SLOs, cost considerations remain. Although calculating value-added estimates themselves is not expensive, the infrastructure they rely on can be costly. For example, districts must work carefully to ensure that their data collection systems are up-to-date and accurate. For a data system to be used for high-stakes purposes such as teacher evaluation, it must meet levels of accuracy not currently found in most districts. If teachers’ estimates are based on students who are not in their classes or on inaccurate student test scores, the system will not be effective.

More broadly, new systems are more complex and higher stakes that those of the past. As a result, districts require more complex and reliable systems for collecting, storing, and analyzing teacher evaluation data. In
early research on the implementation of new systems, the management of evaluation data has been a substantial challenge, especially in smaller, rural, and lower-resourced districts without a technological infrastructure on which to build or personnel to devote to the tasks associated with building and managing a new data system (Donaldson et al., 2014; Firestone et al., 2013).

**Delivery of Evaluation**

Policymakers must also make key decisions about who will carry out the work of evaluation. As with the evaluation measures selected, decisions about who evaluates teachers have implications for an evaluation system’s success in holding teachers accountable or promoting development. Despite clear reasons for using school administrators to evaluate teachers, many principals struggle to find the time or lack the expertise to do so well (Peterson, 2000; Halverson & Clifford, 2006; Johnson et al., 2010). New evaluation systems generally demand even more time from administrators than did traditional systems. In interviews and surveys, teachers and administrators routinely express doubts about administrators’ evaluative skills and knowledge, particularly at the secondary level (Donaldson et al., 2014; Firestone et al., 2013; Hill & Grossman, 2013). And, administrators’ will to evaluate teachers accurately may also be compromised by their need to garner teachers’ trust (Kimball & Milanowski, 2009; Sartain et al., 2011).

Partially in response to these critiques of the traditional model’s reliance on administrators to conduct evaluation, several districts across the country have experimented successfully with identifying expert peers to serve as evaluators (Johnson et al., 2009; Kane et al., 2011). In these systems, expert teachers assume primary responsibility for evaluating a set of teachers in their caseload. Many of these systems are part of a PAR model, where teachers not only evaluate but also support and make employment recommendations for their nontenured and struggling tenured colleagues. Started in Toledo, Ohio in 1981, PAR is now used in approximately 50 districts across the country, and both national teachers’ unions have supported its expansion. PAR has the added benefit of bringing teachers’ unions to the table from the outset of evaluation, thereby potentially reducing union opposition to a teacher’s dismissal (Johnson et al., 2009; Papay & Johnson, 2012). Advocates further argue that peer evaluators may assess teachers more accurately, provide better feedback, and recommend teacher dismissal more often than school principals, perhaps because they have more time and training to evaluate their peers. Of course, these programs also cost more as evaluators must be trained and the intensive observation and support requires a substantial time investment.

**Outcomes of Current Teacher Evaluation Systems**

Given the rapid changes in teacher evaluation, policymakers increasingly demand information about whether evaluation as a reform can improve teaching practice and student outcomes. A growing body of research suggests that rigorous evaluation systems improve teacher practice and raise student test scores, with strong evidence that participating in evaluation can improve teacher skills in a lasting way. At the same time, they spread out the distribution of ratings of teacher performance and lead to more teachers being dismissed or counseled out of the profession. In this way, research suggests that new systems may support both the accountability and development purposes of evaluation.

Research on new evaluation systems encounters many challenges. Districts and states that have implemented evaluation reform have often enacted other, contemporaneous reforms, such as teacher pay or whole-school improvement initiatives. Findings are also context dependent. In the absence of random assignment of district and school personnel, it is difficult to understand whether effects of such systems are due to the selection of particular educators to these settings. Finally, with minor exception, these systems are quite new, so they are only beginning to be studied and researchers have only been able to explore short-term outcomes. Despite these issues, a small but growing body of high-quality studies examines the effects of evaluation.

Evidence suggests that enhanced evaluation systems have spread out the historically skewed distribution of evaluation ratings and led to an increase in teacher departures from districts for performance reasons. For
example, in 2011–12, after evaluation reform in Washington, DC, 16 percent of the city’s teachers received the highest rating and 15 percent received a rating in one of the lowest two categories (Dee & Wyckoff, 2013). In the New Haven (CT) Public Schools, a small fraction—between 2 and 4 percent—of teachers have received the top evaluation rating in the three years of TEVAL’s existence (Donaldson & Papay, 2014). Evidence also suggests that principals do indeed incorporate information from the evaluation system into personnel decisions. In New York City, researchers found that providing information to principals about their teachers’ value-added scores led more low-performing teachers to leave the schools (Rockoff et al., 2012). In Washington, DC, nearly 4 percent of teachers on average were dismissed from the district due to their evaluation ratings, in addition to a substantial increase in voluntary attrition because of low evaluation ratings (Dee & Wyckoff, 2013). Between 1 and 2 percent of teachers in the New Haven (CT) Public Schools have left the district each year due to their evaluation ratings (Donaldson & Papay, 2014). These departure rates are substantially higher than those in earlier years.

In one of the most rigorous studies to date on the impact of district-wide teacher evaluation, Taylor and Tyler (2012) examine Cincinnati’s standards-based evaluation system. At the time, the Cincinnati system had relatively weak incentives for teachers based on their evaluation scores: high scores provided some teachers with an opportunity to apply for leadership positions, and very low scores could contribute to teachers being assigned to the PAR program, with a low probability of dismissal. Despite these weak incentives, the authors found that participating in evaluation improved the performance of mid-career teachers not only during the year they were evaluated, but also in subsequent years. This provides causal evidence that participating in a rigorous evaluation system not only improves teacher performance when teachers are monitored (i.e. undergoing evaluation), but can also build skills that lead to long-term improvement. These improvements are quite substantial (11 percent of a student test-score standard deviation in mathematics). Moreover, the lowest performing teachers prior to evaluation (defined by evaluation scores or value-added scores) experienced the largest performance gains.

More recently, Steinberg and Sartain (forthcoming) present experimental evidence that the implementation of a new evaluation system in Chicago raised student achievement in elementary schools that participated in the first year of the pilot program. After the first year, reading achievement was 0.10 standard deviations higher in the 44 schools that were randomly assigned to implement the new system than in control schools, although math scores were not significantly different. This impact persisted over time. Moreover, achievement gains were larger for treatment schools with higher achievement or lower poverty prior to the experiment. Interestingly, Steinberg and Sartain found that a second cohort of schools did not experience the same increase in student achievement when implementing the new system. The researchers attribute this difference to the fact that training and district support for the second cohort of principals was much less robust than that for the first cohort, who received approximately 50 hours of training in how to observe and provide feedback and weekly support from district leaders.

Finally, studying Washington DC’s IMPACT system, Dee and Wyckoff (2013) also found that the effectiveness of particular teachers improved as a result of participating in the program. Using a regression-discontinuity design, Dee and Wyckoff found that participating in IMPACT led low-performing and high-performing teachers to improve their performance by 0.27 and 0.24 teacher-level standard deviations, respectively. Given their approach, Dee and Wyckoff were not able to estimate an average effect of the program. Because IMPACT combines evaluation with high-stakes consequences such as performance bonuses and dismissal threats, it is difficult to interpret this evidence as purely an effect of evaluation itself. In addition to financial incentives and dismissal threats, IMPACT provides teachers with feedback from classroom observations and with regular coaching to improve their performance. Together, these three studies suggest that teacher evaluation can lead to substantial and lasting improvements in teacher performance, both in districts with strong incentives (e.g., Washington, DC) and those with relatively weak ones (e.g., Cincinnati).

Conclusion

Since 2008, teacher evaluation reform has occurred on an unprecedented scope and scale. While a few districts use the “traditional” model of teacher evaluation held over from the 1980s, many states and districts, spurred
by federal policies such as Race to the Top and No Child Left Behind waiver requirements, have overhauled their evaluation systems, incorporating student achievement and more robust observation data. These sites typically work first to improve the precision of ratings, particularly for observations. Some sites have moved beyond this phase and are beginning to use evaluation data to try to improve instruction. However, states and districts often lack capacity in this regard, and this stage of implementation has been understudied. Emerging evidence suggests that new and enhanced evaluation systems may result in improved teacher performance, as measured by student achievement, and higher rates of selective attrition, thus addressing both of evaluation’s purposes: development and accountability.

Evaluation systems inherently include consequences for teachers, through the simple act of being measured and rated. But, recently, the policy attention to accountability in evaluation has led to increasing calls for states and districts to attach even higher stakes to these ratings, using them explicitly as grounds to dismiss low-performing teachers or to provide monetary rewards to high-performing teachers. For example, Hanushek (2009) has proposed that using evaluation ratings to remove the bottom 5 to 10 percent of the teacher performance distribution would substantially improve student achievement in the United States. In considering such initiatives, it is important to examine the validity of the inferences that these evaluation systems produce and the quality of the teachers who would replace any departing teachers.

Several important questions arise here. First, the research attention to the bias and reliability of teacher performance measures is important, as valid inferences are not possible without good measures. In many places, information on the properties of these measures is absent, and simply because a test is valid and reliable for measuring student achievement in a specific content domain does not mean that it meets these standards for assessing teacher effectiveness. Second, teacher evaluation purports to identify the most “effective” teachers, but no strong consensus exists about what “effective” means (Fenstermacher & Richardson, 2005). To the extent that different teachers have different competencies that contribute to the ultimate goals of schooling, designers of evaluation systems will continue to struggle with how best to summarize a teacher’s contributions in a single rating. Third, the more policymakers tie substantial rewards and sanctions to evaluation, the stronger incentives teachers have to game the system, thereby contaminating the inferences about effectiveness that policymakers hope to draw. For example, teachers may be able to raise their evaluation scores in a variety of ways that do not promote student learning, such as teaching to the test, narrowing the curriculum, currying influence with evaluators, or cheating (Figlio & Getzler, 2006; Jacob & Levitt, 2003; Koretz, 2008). Here, the specific behaviors depend on both the types of measures, as some are likely easier to game than others, and the strength of the accountability consequences attached to the measures. Finally, Rothstein (2013) notes that strong accountability pressures, particularly when they are based on measures that are noisy and unpopular, can affect the career decisions of potential and current teachers.

These considerations do not mean that policymakers should avoid using evaluation for accountability purposes. In fact, there is widespread support, including among teachers and teachers unions, for removing from the profession teachers who remain ineffective instructors even after receiving substantial support (e.g., Weingarten, 2010). Pathways out of the classroom must be more streamlined than they have been in the past, but not so arbitrary that teachers (or potential teachers) worry that employment decisions are based on capricious decisions or evidence that is not valid. The challenge, of course, comes in designing reasonable accountability measures that reflect what the evaluation measures can—and cannot—tell us about teacher effectiveness.

In addition, accountability in isolation is unlikely to promote widespread improvement in teacher quality. Instead, accountability must be coupled with support and opportunities for teachers to learn. Teacher evaluation is a prime policy lever through which to couple accountability and support, as the information embedded in these systems can provide valuable feedback to teachers and help structure individualized learning opportunities. Participating in a robust evaluation system, even one without particularly high stakes for teachers, can improve an individual teacher’s effectiveness over time (e.g., Taylor & Tyler, 2012), suggesting that teacher evaluation’s development pathway is likely a promising approach to improving teacher performance at scale. States and districts recognize this opportunity in rhetoric, but, to date, few have made progress in designing new evaluation systems to promote instructional improvement. This mirrors the relative lack of research on this dimension of evaluation, and highlights an important area of focus for research, policy, and practice in the next decade.
Using evaluation as a system for teacher development requires several steps in implementation. First, evaluation systems must provide feedback to teachers about their practice. Hence, standards-based observations or other measures that can provide rich information that teachers understand and can act upon must remain central to these systems. Second, teachers must buy in to the system and see the standards of instructional practice (and the measures that result) as reflective of effective teaching. Third, evaluators need to be able to provide feedback and school systems must design processes to help teachers use this feedback to improve their instruction. Achieving these ends requires training, time, and the will on the part of district and school leaders to prioritize this work over a multitude of other tasks.

At the same time, research suggests that context matters in the implementation and outcomes of new teacher evaluation models. Implementation considerations such as the purposes of evaluation, the quality and quantity of professional development accompanying new systems, and the extent to which teachers are involved in developing these systems affect their ultimate quality. States and districts cannot simply adopt models that have succeeded elsewhere. And, teachers must be actively involved in the development and implementation of evaluation systems if they are to view them as valuable.

In addition to raising these important policy considerations, our review has identified several topics within teacher evaluation policy that remain understudied. For example, teacher evaluation policy is more centralized today than ever before. What are the implications of this new balance of power at the local, state, and federal levels? Moreover, new teacher evaluation systems are more comprehensive and ambitious than those of the past. We have some understanding of the benefits of these systems, but we have very little information regarding their costs. How do the costs compare to the benefits? As systems seek to incorporate multiple measures of teacher effectiveness, we also need to know more about how best to combine these measures, not simply to improve reliability or limit bias, but to achieve the goals of the system. Finally, while limited research has focused on how best to use evaluation as a system of teacher development, even less attention has been paid to evaluation as a tool for organizational improvement across a school. Most research on evaluation has narrowly attended to the individual, but students in schools move across classrooms, subjects, and grades. As a result, we need better research on how evaluation can work as a human capital strategy for building effective organizations at scale.

Acknowledgments

We thank Jennifer King Rice, Helen Ladd, and Margaret Goertz for their excellent feedback on this chapter. We are also grateful to Susan Moore Johnson, Matthew Kraft, and Eric Taylor for insightful comments on early drafts. Lastly, we thank Daniel Ochs for first-rate research assistance.

Notes

1. Teacher evaluation can also serve as a system to promote organizational performance (e.g., schoolwide effectiveness). However, to date the policy and research world has largely viewed teacher evaluation as an activity to assess and improve individual teacher effectiveness.

2. Research on performance appraisal in economics and organizational psychology highlights important challenges for teacher evaluation systems. In economics, extensions to the principal-agent model describe situations, as in education, where productivity is multidimensional and difficult to measure (Holmstrom & Milgrom, 1991). Some analysts have called for subjective performance measures to provide a more holistic picture of employee productivity. While this approach can resolve some of the multitasking challenges, it can be subject to administrative favoritism and abuse (Baker, Gibbons, & Murphy, 1994). Organizational psychologists have examined such issues, focusing on what Landy and Farr (1980) described as the “social-cognitive process” behind performance appraisal.

3. With the rich performance data embedded in district datasets, more detailed analysis would theoretically be possible. For example, teachers could receive information about their effectiveness with different types of students or with different domains of content on the test. Unfortunately, no research exists about such practices.
References


Brookings Institution.


School Accountability and Student Achievement

DAVID N. FIGLIO AND HELEN F. LADD

Introduction

Demands for more accountability and results-based incentive systems in K–12 education come from many directions and have dominated much of the education policy discussion at both the state and federal levels in the United States for many years (Ladd, 1996; Ladd & Hansen, 1999; Peterson & West, 2003). Accountability in education is a broad concept that could be addressed in many different ways: using political processes to assure democratic accountability, introducing market-based reforms to increase accountability to parents and children, developing peer-based accountability systems to increase the professional accountability of teachers, or using administrative accountability systems designed to drive the system toward higher student achievement. This chapter focuses on this last approach and pays particular attention to programs that use the individual school as the primary unit of accountability.

School accountability in the United States operates largely within the traditional public school system and relies heavily on student testing (Elmore, Abelmann, & Fuhrman, 1996; Clotfelter & Ladd, 1996; Hanushek & Raymond, 2003). Most emblematic is the federal No Child Left Behind Act (NCLB), which became law in 2002. NCLB required states to test students in reading and mathematics in grades three through eight, as well as in one high school grade. In addition, it required states to assess schools on the basis of whether their students (both in the aggregate and by subgroup) were making adequate yearly progress (AYP) toward the ultimate goal of 100 percent proficiency by 2014, and it imposed consequences on schools and districts that failed to make AYP. This law is the most recent incarnation of a bipartisan standards-based reform movement that emerged from a historic 1989 summit in Charlottesville, Virginia, between President George H. W. Bush and the state governors. That meeting generated a set of national education goals that were subsequently embedded in the Clinton administration’s 1994 Goals 2000: Educate America Act. NCLB differs from that act by its far heavier emphasis on accountability and its significantly greater federal intrusion into the operations of individual schools and districts.

School-based accountability programs preceded NCLB in many states. As of 2001, 45 states published report cards on schools, and 27 of them rated schools or identified low-performing ones (Education Week, Quality Counts, 2001). Several states (e.g., North and South Carolina, Texas, Kentucky, and Florida), as well as districts such as Chicago, Dallas, and Charlotte-Mecklenburg, also had full school-based accountability programs in which they rated schools based on their students’ performance, provided rewards either to schools or to teachers for improved performance, and provided some combination of sanctions and assistance to low-performing schools.¹

The United States has not been alone in using student test scores for school accountability. Indeed, centralized reporting of schoolwide examination scores has occurred for over two decades in the United Kingdom (Burgess et al., 2005) and in Chile (Mizala, Romaguera, & Urquiola, 2007). According to 2011 OECD data, 12 OECD or other G20 countries reported school level performance information on national assessments at the lower secondary level in 2009, although the extent to which these countries used that information for high-stakes accountability is not clear.

According to the OECD, many other countries, such as France, Germany, South Korea, and the Netherlands, rely heavily on inspectorate systems to evaluate schools. Some countries, such as England, use both approaches. In inspection systems, professionally trained external review teams visit individual schools on a periodic basis and write public reports highlighting the strengths and weaknesses of the school. Some...
inspection systems are intended to be entirely evaluative, while others take the next step of advising schools how to improve (Ladd, 2010). Most of this chapter focuses on the U.S. experience with test-based accountability, albeit with some limited attention to inspection systems.

The Rationale for School-Based Accountability

The current school-based accountability movement emerged out of a desire, particularly in the United States and the United Kingdom beginning in the 1980s in the Reagan and Thatcher eras, to measure performance in the public and nonprofit sectors (Figlio & Kenny, 2009). In the United States, this effort aligned well with the broader standards-based reform movement in education (O’Day & Smith, 1993). That movement involved the setting of clear, measurable, and ambitious performance standards for students across a number of core subject areas, aligning curricula to these standards, and establishing high expectations for students to meet them. Assessment of students is a key component of standards-based reform. These assessments are used to measure student progress toward mastery of the standards and also the effectiveness of the schools they attend.

In the context of standards-based reform, accountability is only one part of a larger policy package. Accountability—whether for schools, students, teachers or districts—can also be viewed as a stand-alone policy. One rationale for such a policy comes from comparisons to the private sector where business firms focus attention on results and rely on benchmarking procedures to measure progress. Another rationale is provided by the economists’ model of the principal-agent problem. According to this model, school administrators and teachers may underperform because state policymakers do not have a good means of monitoring them. It follows that student achievement would improve if state policymakers could monitor the teachers and school administrators more effectively.

The information content in school accountability systems can potentially provide a powerful mechanism for overcoming the principal-agent problem. Assessing schools against the common metric of standardized student test scores provides policymakers and members of the general public with independent information regarding how well schools and school districts (and potentially teachers) are doing in comparison to their peers or to external performance standards. At the same time most administrative accountability systems measure only a very small fraction of the educational outcomes that stakeholders value. The outcomes in such systems typically include student achievement as measured by test scores in only the core subjects of math and reading, supplemented in some cases (as under NCLB) with some nontest-based measures such as student attendance or graduation rates. As documented by Jacobsen & Rothstein (this volume), educational stakeholders value a much broader range of academic outcomes as well as other outcomes such as citizenship. Thus, the narrow focus of test-based administrative accountability systems on test results in math and reading clearly privileges one narrow set of outcomes over others.

In addition to promoting more effective use of resources toward specified outcome goals, school accountability systems can also be designed to promote equity by shining the light on disparities in inputs and outcomes for groups such as ethnic minorities and students with disabilities. Indeed, the name “No Child Left Behind” makes clear that the federal accountability system in the United States was motivated at least to some degree by equity concerns, and its twin pillars of promoting equity and higher achievement are potential reasons for why the law enjoyed such strong bipartisan support at its outset.

Design Matters

How a school accountability system is designed can have a significant impact on the nature of and the strength of the incentives that schools face to raise student achievement in the tested subjects. Moreover, the design can affect which students receive the most attention.

The Consequences of Accountability
The consequences—and hence incentives to improve—associated with accountability can take two different forms: explicit positive rewards and/or negative sanctions built into the system or as pressure from external stakeholders that stems from the information provided by the accountability reports. Positive consequences for high-performing schools may include increased resources, greater school level autonomy to spend these resources, or teacher bonuses. Possible sanctions for schools not meeting performance expectations include the withdrawal of autonomy, requirements for how the school must respond, school restructuring or, in the extreme, closure of the school. Several studies, including Hanushek and Raymond (2005) and Dee and Jacob (2011), specifically identify systems with “consequential” accountability and provide evidence that the inclusion of specific consequences appears to induce educators to improve student outcomes.

Another form of pressure on school comes from parents and community members. The broader economics literature on the role of information on product quality (e.g., Jin & Leslie, 2003; Mathios, 2000) shows the power of information to affect behavior. Hence, the increased availability of information about specific schools is likely to generate community and local pressure on schools. Studies showing that differences in school test scores across school district boundaries are capitalized into housing values confirm that families care about school level outcomes (Black, 1999). Figlio and Lucas (2004) go one step further and demonstrate that housing markets clearly react to the information embedded in school accountability systems. The findings of these two studies have since been replicated in numerous settings in North America and Europe. Accountability reports also affect the choices parents make about individual schools (see Figlio & Lucas 2004; Hart & Figlio, 2014; and Hastings & Weinstein, 2008). Finally, Figlio and Kenny (2009) document that parents and community members withhold private contributions from schools that are reported to be performing poorly and offer more financial support to those that are performing well. Pressures of this type thus provide an additional mechanism through which accountability systems may influence educator behavior. One implication is that accountability systems may induce change even when the explicit rewards or sanctions they include are quite small or, alternatively, so large as not to be credible.

Scope and Domains of Accountability Indicators

Policymakers face important trade-offs when determining how broad-based to make their accountability systems. On the one hand, holding schools accountable for a small set of outcomes provides incentives for schools to narrow the scope of the education they provide. On the other hand, a broad set of outcomes is more difficult to measure reliably and may blur the focus of school and district personnel.

U.S. states differ in the substantive breadth of their state accountability systems. While most states assess schools primarily on the basis of reading, mathematics, and sometimes writing, some states implement a broader approach. Virginia, for example, tests students in additional subjects, including science, U.S. and Virginia history, and social studies, and reports on end-of-course examinations in high school. Kentucky’s core content areas for school accountability also include fine arts and humanities. Until recently, Nebraska assessed schools on a narrow set of substantive areas but at greater depth by including portfolio reviews as part of the assessment.

The available evidence strongly supports the conclusion that schools tend to concentrate their attention on the subjects tested and on the grades that have high-stakes tests (Deere & Strayer, 2001; Ladd & Zelli, 2002; Stecher et al., 2000). Other studies (e.g., Hamilton et al., 2005; Jones et al., 1999; Koretz & Hamilton, 2003; Linn, 2000; Stecher et al., 1998; and Stecher et al., 2000) show that teachers and schools tend to narrow the curriculum and shift their instructional emphasis from nontested to tested subjects, while earlier work by Shepard and Dougherty (1991) and Romberg et al. (1989) suggest that teachers focus more on tested content areas within specific subjects. In related work, Chakrabarti (2005) presents evidence that schools may concentrate their energies on the most easily improved areas of instruction, rather than on all aspects of the tested subjects.

However desirable it may be, increasing the scope of testing is costly, in terms of both financial costs and the opportunity cost of foregone instructional time that is devoted to testing. Broadening the scope is also limited by the technology available to assess student progress because some subjects are simply more challenging to assess than are others. Given the well-established tendency of educators to focus their attention on the material most likely to be covered on the assessment, a behavior known commonly as “teaching to the test,” it seems likely that educators will concentrate on the assessed components of difficult-to-assess subjects rather than on
the components that may be more highly valued but more difficult to measure. A recent National Research
Council panel (Wilson & Bertenthal, 2006) warns of this potential with regard to science assessment, as the
members note how challenging it is to design a science assessment to test students’ scientific inquiry skills.

Outcomes-based accountability systems can be modified to incorporate various other measures of school
performance besides student test scores. Hanushek and Raymond (2003) construct a hierarchy of nonetest
indicators of school performance, ranked on the basis of their relevance and likely alignment to objective
measures of school progress. For instance, they argue that certain measures such as the dropout rate,
graduation rate, number of students in advanced courses, percent of students passing end-of-course exams,
retention rate, student mobility, and suspension rate, are relatively closely related to student achievement and
could, therefore, be incorporated into accountability systems designed to raise student achievement. One
shortcoming of using any of these broader outcomes, however, is that school officials may find many of them
relatively easy to manipulate, thereby circumventing the goals of the accountability system.

Considerations of this type, along with more fundamental concerns about the extent to which it makes sense
to hold schools accountable for student outcomes that are determined not just by schools but also by nonschool
factors, such as the family backgrounds of the students, argue in favor of using inspection systems rather than
outcomes-based accountability (Ladd, 2007 and 2012). With their focus on internal school practices and
processes, inspection systems can provide a broad holistic assessment of school organization, curriculum,
instructional policies and classroom practices, and need not be limited to the very specific sets of skills and
competencies that are most easily measured by student assessments and other achievement related outcomes.

Measuring School Performance Using Standardized Tests

Policymakers have relied on two main approaches to measure school performance using student test scores.
"Status" measures judge a school’s performance based on levels of performance, such as the fraction of students
attaining a given proficiency level or the average test score in the school. "Growth" measures, often called
"gain scores" or "value-added" measures, evaluate schools on the degree to which their students improve in
their test performance from one year to the next (or, in some cases, from fall to spring of a given school year).

Growth measures can be technically complicated, and a thorough discussion of the issues is beyond the scope
of this chapter. The simplest of these measures averages year-to-year changes in test scores across all students
in a school while more complicated measures use regression models to adjust changes in test scores for various
student characteristics or to take into account the variance in observed changes. NCLB is based on a “status”
model of evaluating schools, though the U.S. Department of Education has recently granted some states the
authority to supplement status measures with growth components. Many state-level systems, as well as
proposals for overhauling federal accountability, further expand the use of growth models in the United States.

Because accountability systems based on status and growth measure different outcomes, they tend to
generate different incentives for schools. Status-based systems that focus on the percent of students who
achieve at proficient levels seek to encourage schools to raise performance at least to that level (Krieg, 2008;
Neal & Schanzenbach, 2010). This approach is appealing to many policymakers because it sets the same target
for all groups of students and encourages schools to focus attention on the set of low-performing students who
in the past may have received little attention. Status-based systems also have the advantage of being
transparent.

The goal of the growth-model approach is to encourage schools to improve the performance of their students
independently of the absolute level of that achievement. Such an approach is appealing to many people because
of its perceived fairness. It explicitly takes into account the fact that where students end up is heavily
dependent on where they start, which tends to be highly correlated with family background characteristics. At
the same time, the use of the growth-model approach may raise political concerns, both because the public
may find the approach less transparent than the status approach and because some see it as a way of letting
schools with low average performance off the hook.

Systems using status and growth models generate different incentives in part because they lead to different
rankings of schools. Many schools deemed ineffective based on their aggregate performance levels may
actually have quite high "value added" and vice versa (Clotfelter & Ladd, 1996; Ladd & Walsh, 2002; Kane &
Staiger, 2002; and Stiefel et al., 2005). Some state-specific accountability systems (e.g., Florida and North
Carolina) encourage both high levels of performance and high test-score growth by including both levels and gains in the index of school success (Ladd & Zelli, 2002). With all states subject to the status requirements of NCLB, schools in some states are subject to accountability under both approaches and sometimes face conflicting signals about their relative success or failure.

The status and growth approaches also send different signals to schools about which students deserve additional attention. Under a status-based system designed to encourage schools to raise student performance to some threshold level, the position of the threshold matters. A challenging performance threshold—one that would be consistent with the high aspirations of the standards-based reform movement, for example—would provide incentives for schools to focus attention on a larger group of students than would be the case with a lower threshold. Evaluating schools on the basis of “value added,” by contrast, provides incentive for schools to distribute their effort more broadly across the entire student body. In such a system, however, schools may have an incentive to focus attention on the more-advantaged students if the test score gains of those students are easier to increase, bringing up the average gains for the school (Ladd & Walsh, 2002; Richards & Sheu, 1992; Ladd & Lauen, 2010).

Under either approach, random errors in the measurement of student performance can generate inconsistent rankings of schools over time—a factor that weakens incentives for improvement. Measurement error is a particular problem for small schools because the fewer the students there are in a school, the larger is the schoolwide average measurement error, and hence the less consistent the school’s ranking is likely to be from one year to the next. As a result, schools deemed to be improving at one point in time may be judged as declining the next year simply because of measurement error (Kane & Staiger, 2002). Measurement error is exacerbated when schools are rated based on the growth model because it requires student test scores for two years, both of which are measured with error. When school personnel receive inconsistent signals from one year to the next, they have little incentive to respond in a constructive way. The policy issue is whether the error is large enough to mask the signal that drives the incentives in the accountability system.

Neither the status nor the growth approach to measuring school performance perfectly captures school efficiency—the effectiveness with which schools use their resources to maximize student outcomes given the students they serve. Because efficiency cannot be observed directly, it must be inferred from statistical analysis that controls both for the resources available to the school and the characteristics of the school’s students (Stiefel et al., 2005). If the goal of an accountability system is to induce schools to use the resources they have more effectively, then, in principle, schools should be rated on their efficiency, not simply on the level or growth of their students’ achievement. The problem is that the data requirements for such efficiency measures are often daunting and the statistical techniques can be complex (Ladd & Walsh, 2002; Stiefel et al., 2005).

In contrast to a measure of school efficiency, the status and growth measures provide information on whether or not schools are meeting expectations for either the level of achievement or the growth in achievement with no attention to what accounts for that performance. Although inefficient use of available resources may be one reason for poor performance, another could well be that the resources available to the school are insufficient for the school to meet the accountability standard given the profile of the students in the school. In the latter case, it is neither fair nor likely to be productive for state or federal policymakers to hold the teachers or other school personnel responsible for the poor performance of the school’s students (Ladd & Walsh, 2002). Thus, accountability and the financing of schools are closely intertwined.

**Exclusions**

Designers of accountability systems must also determine which students should be counted when evaluating the contribution of a school to student learning. While it may seem clear that all of a school’s students should be credited to the school—especially when an accountability law has a name such as No Child Left Behind—universal inclusion raises important questions about fairness and attribution. For instance, should a school be held responsible for the performance of a student who arrived at the school shortly before the test was administered? Should schools be held responsible for students for whom testing is more challenging, or potentially less reliable, such as students with disabilities? Florida’s treatment of mobile students in successive iterations of its accountability system illustrates the importance of this issue. When Florida introduced its school rating system in 1999, the state included in its calculations of school grades both the students who spent
the full academic year and those who were recent in-migrants. The following year the state amended its policies to include only students who had spent the full academic year, and these rule changes changed the sets of schools identified as low- or high-quality (Figlio & Lucas, 2004). The key question becomes whether the system is setting reasonable expectations for educators in specific schools.

Similar challenges are present with regard to other types of exclusion decisions, such as whether to exclude students with disabilities and/or language minority students. Policymakers face clear trade-offs with respect to the treatment of these special populations. On the one hand, schools with large fractions of mobile and disabled students in many cases have a legitimate argument that holding them accountable for the academic achievement of such challenging-to-educate students puts them at an unfair disadvantage relative to other schools with fewer disabled students. On the other hand, excluding students on the basis of classification provides schools with less incentive to support these students as well as an incentive to selectively reclassify or move students in order to look better against performance metrics. The evidence is quite clear that schools have responded to accountability pressures by reclassifying low-performing students as students with disabilities (see, e.g., Cullen & Reback, 2006; Deere & Strayer, 2001; Figlio & Getzler, 2007; and Jacob, 2005). In general, these incentives are smaller under the growth approach than under the status approach.

Why Accountability Might not Improve Student Achievement

The preceding discussion predicts that school accountability programs will increase student achievement, although the magnitude of the predicted effects for particular groups of students or types of schools may well differ depending on how the system is designed. For several reasons, however, school accountability systems might not generate higher achievement.

Improving Measured, But Not Generalizable, Achievement

Monitoring provides incentives for those being monitored to appear as effective as possible against the metric being used. It is certainly possible, therefore, that educators could teach very narrowly to the specific material covered on the tests, with little or no generalizable learning outside of that covered on the test (Koretz & Barron, 1998). The magnitude of this concern depends largely on the extent to which the tests cover the full range of valued material; the more limited is the covered range, the greater is the distortion caused by teaching to the test. In addition, though, some educators may go further than teaching to the content of the tests by devoting substantial time to test-taking strategies that have little long-term benefit for students. Some may even engage in outright cheating to make their school look better. Jacob and Levitt (2003) document, for example, that a small fraction of Chicago teachers responded to accountability pressures in that city by fraudulently completing student examinations in an attempt to improve observed student outcomes. Similar findings have been documented for other large cities including Atlanta and Washington, DC.

A common approach for determining the extent to which an accountability policy has resulted in generalized learning involves examining whether gains observed based on high-stakes tests are also observed for low-stakes tests that have no specific consequences for schools or students but at the same time measure valued knowledge and skills. A natural test for that purpose is the National Assessment of Educational Progress (NAEP), which has been administered to a nationally representative random sample of students since the early 1970s and to representative samples of students in grades four and eight in most states since the 1990s and in all states since the introduction of NCLB. Because of its high profile and national scope, the NAEP has been widely employed in the studies described later in this chapter assessing the effects of accountability on student learning.

Although observing performance on a low-stakes test has considerable appeal, it has at least two potential downsides. One is that students may not take a low-stakes test sufficiently seriously to do their best work. Unless student effort differs from one administration of the low-stakes test to the next, however, changes in performance on the low-stakes test should provide a reasonable estimate of gains in student learning. A second downside to using low-stakes tests is that the high-stakes tests are often aligned to the standards valued by
policymakers, while the low-stakes tests are not as well aligned. Hence, findings of smaller effects of accountability when low-stakes tests are used to measure performance may simply reflect differences in the degree to which the two types of tests reflect the material that policymakers expect schools to cover.

Differences in the patterns of results for high- and low-stakes tests are quite common. To anticipate the later discussion of the evidence, we note here the significant differences in findings for Texas in terms of overall test scores and the findings related to subgroups (Klein et al., 2000). The increase in test scores based on the NAEP were about a third of that based on the state test (the Texas Assessment of Academic Skills [TAAS]), and achievement gaps between black and white students increased somewhat based on the NAEP but decreased substantially based on the TAAS. Klein et al. (2000) speculate that the reasons for the differing patterns of results for the two tests is that Texas teachers may be teaching very narrowly to the TAAS and that the schools serving minority students may be doing so even more than other schools.

Additional evidence on whether the transferability of knowledge from high- to low-stakes tests emerges from Jacob’s 2005 study of accountability in Chicago. Jacob compared achievement gains for fourth and eighth graders in math as measured by scores on the district’s high-stakes test to those on a comparable, but low-stakes, test administered by the state of Illinois. Those comparisons show that gains for eighth graders generalized to the state test but that those for fourth graders did not. In Florida, Figlio and Rouse (2006) find consistently smaller estimated effects of accountability on low-stakes tests than they do using high-stakes tests. That said, recent findings by Craig et al. (2013), Reback et al. (forthcoming), and Rouse et al. (2013) demonstrate that measured improvements in test scores associated with accountability pressure are accompanied at least in part by substantive changes in instructional policies and practice and resource allocation.

**Manipulating the Test Pool and Other Strategic Behavior**

Teaching to the test is not the only mechanism through which schools might alter their behaviors in response to the incentives embedded within accountability systems. Considerable evidence documents that schools engage in strategies that artificially improve test scores by changing the group of students subject to the test. The most widely studied behavior of this type is the selective assignment of students to special education programs. As mentioned above, many studies show that schools tend to classify low-achievers as learning disabled in the context of accountability systems. Although the greater rates of classification may not be undesirable in all cases, they highlight the possibility that schools are manipulating the testing pool specifically to inflate measured school performance. Moreover some of these decisions may have spillover consequences outside of education. Bokhari and Schneider’s (2009) finding that school accountability policies affect the diagnosis and medication for ADHD, for example, highlights the possibility of negative health consequences. Other evidence of strategic behavior designed to artificially improve test scores by manipulating the pool of test takers comes from Figlio (2006), finding that some Florida schools changed their discipline and suspension patterns around the time of the testing in ways consistent with the goal of improving test-takers’ average scores, and Ozek (2012) finding that students who are not counted for accountability purposes perform worse than those who are.

Evidence suggests that schools also engage in other types of strategic behavior designed to raise test scores. For example, Figlio and Winicki (2005) demonstrate that some schools change their meals programs at the time of the tests in an apparent attempt to raise performance on high-stakes examinations. In addition, Boyd et al. (2008) illustrate how high-stakes testing in certain grades in New York altered which teacher taught in particular grades and schools, and Fuller and Ladd (2013) show that elementary-school principals in North Carolina responded to the test-based pressures of NCLB by moving their weaker teachers away from the tested grades toward untested earlier grades, to the detriment of young learners.

Many of these behaviors are less likely to occur in growth-model than in status-based systems. The reason is that in the growth approach, the manipulative behavior that increases student achievement in one year would make it more difficult for the school to attain accountability goals the following year. No such trade-off arises in status-based accountability systems.

In a federal system with multiple levels of decision making, subfederal units may themselves respond strategically to federally imposed accountability pressures in ways antithetical to higher achievement. For example, under NCLB states have incentives to set low proficiency standards. That is the case because NCLB...
delegates to states the task of defining proficiency standards at the same time that it imposes penalties and sanctions on schools and districts that fail to make adequate progress toward those standards. Peterson and Hess (2006) document the low level of concordance between the progress of a state’s students toward the state-defined proficiency standard and their performance on the NAEP. States such as South Carolina that set very high standards for their students find themselves with large fractions of schools deemed in need of improvement, while states such as Texas that set low standards have few such schools. This interaction between state-set standards and the likelihood that their schools will face sanctions has the potential to lead, in Peterson and Hess’s words, to a “race to the bottom” in terms of setting proficiency standards.

**Lack of Capacity to Respond Productively**

While the evidence clearly shows that educators respond to incentives, they do not always respond in ways consistent with the intended goals. One reason may be that external incentives may be too small to override the professional judgments of teachers and school administrators. Indeed, Bacolod, DiNardo, and Jacobson (2009) suggest that schools that receive additional accountability bonuses in California do not spend these bonuses in ways that improve student outcomes. Additionally, as Frey (2000) points out, if the extrinsic incentives associated with the accountability system crowd out the intrinsic motivations that attracted educators into teaching and made them effective, one might observe stagnant or decreasing performance by students in response to a new accountability regime.

School personnel also might fail to respond to incentives embedded within accountability systems if they lack the capacity to respond in ways desired by state or federal policymakers. Some schools may have insufficient resources to effect serious change in student outcomes, and others may lack the leadership required for significant change. Diamond and Spillane’s (2004) case studies show that higher-performing schools were able to focus on enhanced performance in ways that lower-performing schools could not. More generally, teachers may lack the necessary skills and knowledge to meet the expectations of an accountability system that expects rates of improvement far larger than historical experience has shown to be feasible, as was required under the initial iteration of NCLB. Thus, it could be that one of the major assumptions underlying stand-alone accountability programs—namely that teachers and schools are underperforming because of insufficient monitoring of their behavior—is incorrect. If school resources must be at a certain level to bring about positive performance improvements, or if principals and teachers have sufficient resources but lack the specific policy and practice knowledge necessary to implement highly successful instructional policy and practice changes, then accountability might not lead to meaningful improvements in student outcomes.

The lack of potential responsiveness to accountability could be exacerbated by the fact that accountability systems generally concentrate on shorter-term achievement improvements, while many of the policies and practices that schools may wish to implement can take longer to bring to fruition. School accountability may solve one principal-agent problem by introducing a new one—educators may eschew the types of policies that might yield large-scale long-term success in favor of those that would be less successful in the longer term but might generate bigger boosts today.

**Summary**

Despite the theoretical prediction that school accountability systems will improve student achievement—at least for certain segments of the school population—such gains are not a foregone conclusion. In some cases schools may focus on test scores to the exclusion of transferable knowledge or engage in strategic behavior that is harmful to achievement. Potentially most important, schools may lack the knowledge and capacity to produce significant gains in student achievement.

**The Effects of Accountability on Student Achievement**

Measuring the effects of test-based accountability systems on student achievement is not a simple task. When
such systems are part of a larger standards-based reform effort, it is difficult to separate the effects of the accountability system from those of other components of the reform package. In addition, researchers face the challenge of finding appropriate control groups to determine what would have happened to student achievement in the absence of the accountability system. That is a particular challenge for a national program such as NCLB, and it has forced researchers to use a variety of creative approaches.

Dee and Jacob (2011) address the challenge of measuring the effects of NCLB by comparing achievement trends in states that had no school-level accountability systems prior to NCLB to trends in states that did, on the grounds that the introduction of NCLB placed new accountability pressure on the former states but essentially continued the existing pressure in the states that already had accountability. They conclude that NCLB induced positive and statistically significant gains in NAEP scores in the affected states for fourth graders in math, and positive, but not significant, gains for eighth graders in math. They find no effects for test scores in reading. Also using NAEP data, Wong, Cook, and Steiner (2009) find positive effects of NCLB in fourth- and eighth-grade math when they compare the change in achievement in Catholic schools, which were not subject to NCLB, to the change in public schools since implementation, and when they compare the growth in states with low proficiency standards, and thus with fewer schools failing to meet NCLB goals, to those with higher proficiency standards. In contrast, Lee and Reeves’s (2012) study based on an interrupted time series design finds less positive effects of NCLB. Finally, using a research strategy based on longitudinal data on individual students, Cronin et al. (2005) compare the scores of more than 320,000 students across 23 states based on tests administered by the Northwest Evaluation Association in the year just before NCLB was implemented (2001–02) to those in the year just after it was implemented (2003–04). Although they find higher achievement levels post NCLB, especially on the math tests, they also report lower rates of growth in test scores from fall to spring following accountability.

Focusing on the distributional effects of NCLB, Neal and Schanzenbach (2010) compare test scores of students who took the same test right before and right after NCLB was passed and found substantially higher post-NCLB scores among students in the middle of the achievement distribution in Chicago. Using North Carolina data, Lauen and Gaddis (2012) suggest that the effects of NCLB’s subgroup requirement may be uneven, with positive effects for minority and disadvantaged students, but potentially adverse effects for very low and high achievers in mathematics, a result consistent with those of Neal and Schanzenbach. Taken together, these studies provide some evidence that NCLB increased student test performance, especially in math, but with uneven effects across the distribution of students.

Many of these NCLB studies build on an earlier accountability literature that compared states that implemented state level accountability systems introduced well before NCLB to those that did not, as well as early versus late adopters of accountability (Amrein & Berliner, 2002; Braun, 2004; Carnoy & Loeb, 2002 and 2005; Hanushek & Raymond, 2005; Rosenshine, 2003). Lee’s (2006) meta-analysis of 12 cross-state studies completed between 1994 and 2004 indicates that the adoption of accountability at the state level led to a small improvement in student test scores.

In addition, a number of early studies examined the effects of specific state or local accountability systems. The most compelling of these studies have the advantage of making use of programmatic rules to help to identify the effects of accountability on student outcomes. The main advantage of district and state specific studies is that the analysis is firmly focused on a specific, well-defined accountability system. Some of the studies, especially those for particular states, however, are hampered by the difficulty of predicting what would have happened to student achievement in the absence of the state’s accountability system, and most of these studies are best thought of as the effects of differential degrees of accountability pressure, rather than global effects of accountability per se.

The findings of these studies are generally positive, but they are decidedly mixed. One set of papers describes trends in student achievement before and after specific states or districts implemented accountability programs. Richards and Sheu (1992) find positive trends in South Carolina following reform in that state. Using a more sophisticated interrupted time series design, Jacob (2005) also finds positive trends in both math and reading scores on the high-stakes tests following the introduction of accountability in the city of Chicago, but the positive results on the district tests show up on the lower-stakes state test only for eighth graders and not for third graders. As we mentioned earlier, Klein et al. (2000) compares high- and low-stakes tests in Texas following reform and finds far more positive results for the state’s high-stakes test than for the lower-stakes tests.
NAEP, although even the low-stakes exam showed positive trends.

A different group of district- or state-specific studies use variation in accountability pressures on specific schools within a given system. Figlio and Rouse (2006), West and Peterson (2006), Rouse et al. (2013), and Chiang (2009) all exploit discontinuities in school accountability grades and find positive effects of receiving low grades on student achievement gains with effects up to 0.20 standard deviations, though most are between 0.05 and 0.10. Rockoff and Turner (2010) take a similar approach in New York City and find positive effects of accountability pressures associated with receiving a failing grade. Since schools on both sides of the failing cutoff face some accountability pressure, these studies are best interpreted as the effects of differential degrees of accountability pressure, rather than systemic effects of accountability. Taken together, the district- and state-specific studies, like the studies of NCLB, provide some evidence of a positive relationship between accountability and student achievement, but they are not universal in this conclusion.  

While, in general, the available studies indicate that accountability pressure may raise student achievement somewhat, positive effects emerge far more clearly and frequently for mathematics than for reading. This pattern is particularly clear when the outcome measure is based on a national test, such as NAEP, but it also emerges in some of the district or state level studies such as Figlio and Rouse (2006). In part this pattern reflects the fact that some authors report results only for math, although that is presumably because of the smaller effects for reading. The larger effects for math are intuitively plausible and are consistent with findings from other policy interventions such as voucher programs (Zimmer & Bettinger, this volume) and tax and expenditure limitations (Downes & Figlio, this volume). Compared to reading skills, math skills are more likely to be learned in the classroom, the curriculum is well defined and sequenced, and there is less opportunity for parents to substitute for what goes on in the classroom (Cronin et al., 2005).

Several studies have documented that school accountability systems have had long-lasting effects on student test scores, even after the students have left the schools directly affected by accountability pressure. Rouse et al. (2013) and Chiang (2009) both show that student test scores in mathematics, and to a lesser degree in reading, are persistently higher for several years following a student’s departure from an affected public school. Deming et al. (2013) find that the gains associated with a Texas school labeled as low performing include improved graduation probabilities and postsecondary attainment and higher earnings at age 25, although the long-term impacts on low-scoring students were negative. This evidence provides support for the view that at least some of the estimated test score responses to school accountability pressure represent genuine gains in learning.

That said, the positive effects of test-based accountability on student test scores and possibly on longer-term outcomes need not imply that test-based accountability is desirable, or the only form of accountability that could generate improved student outcomes. In his study of the English inspectorate system, Hussain (forthcoming) shows that an inspectorate-based approach to school accountability has positive effects on student test scores independent of any benefits of the test-based accountability component of the English system. This finding indicates that if a jurisdiction were to implement an inspection system, the test score gains associated with test-based accountability systems might be realized without the undesirable distortionary behavioral responses that are common in test-based accountability systems. While it is supported by the available evidence, this last point is still speculative and deserves more attention from researchers. An additional and broader argument for greater attention to inspection systems is that they may well prove superior to test-based accountability in terms of their ability to promote a richer set of student skills and dispositions than those measured by test scores, albeit most likely at a higher financial cost (Ladd, 2012).

**Future Research**

While some progress has been made in the period of time between the two iterations of this chapter, it is still the case that very few papers look inside the “black box” to uncover the mechanisms through which accountability policies might affect instructional policy and practice, and student outcomes. Doing so is certainly a tall order: typically we have just one instrument for accountability, but many possible pathways through which it could be operating. That said, any research that can describe the conditions under which accountability might have better or worse outcomes would be highly desirable, as would studies that attempt to identify the factors within an accountability system that make the system more or less conducive to success,
and for which groups of students.

We currently also know extremely little about the longer-run effects of accountability on students. Even if accountability improves student test scores, does this come at a cost of higher-order thinking skills or other types of executive functioning, socio-emotional skills, or the like, or are these other forms of cognitive skills also improved? Do students subject to school accountability do better or worse as adults? And as the teacher labor force changes as a consequence of school (or teacher) accountability systems, do students thrive or suffer to differing degrees? These are the outcomes that matter the most regarding school accountability, and to date we know almost nothing along these lines.

In addition, almost all the existing quantitative evidence on the effects of accountability is based on test-based accountability systems. But as we have highlighted throughout this chapter, there are different institutional structures for school accountability, including inspectorate systems. Studies that can gauge the benefits and costs of inspection systems in addition to or instead of test-based accountability regimes would be extremely valuable in helping accountability-minded governments make well-informed choices about the specific nature of the accountability regime that they wish to implement.

Notes

1. The concentration of early accountability efforts in the South was motivated by the desire of southern governors to foster economic development; the fact that for historical reasons state governments in the South typically had more authority over education finance and governance than in other parts of the country and hence were in a position to impose accountability; and that teachers unions, which might have opposed accountability programs, were not a major factor in most southern states.

2. Data were not available for a comparable analysis of reading scores.

3. Moreover, these results imply that the effects of NCLB occurred nearly immediately and before any sanction regime was put in place.

4. Experimental evaluations of test score reporting at the school level, such as Andrabi et al.’s (2013) new results from Pakistan, also support the notion that accountability can boost student test scores.

5. One exception to this finding of larger effects for math emerges from Jacob (2005), where the positive effects for low-performing students were somewhat stronger in reading than in math.

References


School Competition and Student Outcomes

BRIAN GILL AND KEVIN BOOKER

Introduction

An entire section of this volume is devoted to issues related to school choice and educational privatization, including charter schools, the public funding of private schools, and educational management organizations that operate public schools under contract. Nonetheless, the use of market forces in education is an appropriate topic for this section as well, because market-based educational policies may affect the productive use of resources across the educational system. In particular, many supporters of market-based approaches believe that the competitive pressures they create will promote an increase in school productivity (see, e.g., Friedman, 1955; Chubb & Moe, 1990; Hoxby, 2000). If so, benefits might be seen not only by students enrolled in schools of choice (whether private, charter, or privately managed), but also by students who remain in conventional public schools and by taxpayers. At the same time, the adoption of market mechanisms in schooling raises questions about the extent to which schools operating under a competitive model will productively serve public as well as private purposes (see, e.g., Fiske & Ladd, 2000; Fuller, Elmore, & Orfield, 1996; Gill et al., 2007; Gutmann, 1987; Lake & Miller, 2012; Levin, 2000; Wells, 1993; Wolf & Macedo, 2004). How will competition affect the stratification of students across schools, by race, social class, and ability? Will privately operated schools effectively promote education in the knowledge, skills, and values needed for effective citizenship in a democracy?

This chapter does not address the productivity of market-based schools as measured by the achievement of their own students; we leave that issue to the chapters in Section V of this volume. Instead, we address the effects of competition on conventional public schools and on the traditional public purposes of education, including student integration and the education of citizens.

A variety of different kinds of policies can promote market forces in education, potentially producing competitive effects. Scholarships that subsidize tuition at private schools, usually known as vouchers, have been proposed in a variety of forms for at least 50 years. Publicly funded voucher programs are currently operating in Ohio, Florida, and Milwaukee, Wisconsin, as well as several other states and districts. Some states use state income tax policy to subsidize private-school tuition payments indirectly. This subsidy can involve permitting deductions or credits for families paying private-school tuition or, as Arizona, Pennsylvania, and Florida have done in recent years, giving tax credits to individuals or businesses that contribute to privately operated scholarship/voucher programs. Over the last 20 years, the most popular market-based policy intervention in state legislatures around the country has been the creation of charter schools—publicly funded and publicly regulated schools that are privately operated (i.e., operated autonomously, outside the direct control of local public officials) and enrolled by student/parent choice. Although the first charter schools were established only two decades ago, over 6,000 charter schools are now operating in over 40 states, serving 2.3 million students. Meanwhile, competition might also result from choices among conventional public schools, whether created by intradistrict or interdistrict open enrollment policies or simply arising from families’ options to set up residence in different school districts. Economists typically refer to this latter form of choice as “Tiebout choice” after the economist who first described the process of voting with one’s feet for public services.

Whether competition in the K–12 education market will produce positive results has been a matter of great theoretical debate for half a century, since Milton Friedman (1955) proposed a voucher system. Supporters of competition argue that it will give families a wider range of choices and make schools more responsive to their
customers. This kind of responsiveness should lead to an increase in school productivity, under the assumption that parents and students will choose the most effective schools.

But K–12 education has characteristics that have led skeptics of market solutions to doubt that competition will produce desirable outcomes. Markets are effective tools for generating efficient distributions of private goods. But markets that efficiently allocate private goods often undersupply public goods. For two centuries, a primary justification for the public support of education has been the understanding that an educated citizenry, socialized in common civic values and equipped with the skills to be economically productive, is a public good. Although no one in mainstream politics seriously questions public funding for education, proposals for competition and privatization raise a question about whether public purposes will be served through the pursuit of private ends.

Educational consumers—parents—are presumably generally interested in sending their children to schools that are academically effective. Nonetheless, market failure might occur in the educational marketplace because both schools and parents are often interested not only in the quality of schools, but also in the characteristics of the student body. If schools have the implicit or explicit ability to select their students, or if parents select schools based partly on peer characteristics, then competition could increase stratification by ability, race, and socioeconomic status (see Wells, 1993; Fiske & Ladd, 2000; Epple & Romano, 2000; and Schneider et al., 1997). Moreover, the fact that education is compulsory distinguishes it from most market goods: some parents and students are likely to be passive choosers who will be disadvantaged in a marketplace where others are active choosers. In addition, schools might pursue strategies to make themselves effective selectors of students rather than productive educators. And some parents and schools might care very much about promoting particularistic private values (such as, for example, football or religion) rather than academic effectiveness and civic socialization.

In addition, conventional public schools are not institutionally designed to respond effectively to market pressures (see, e.g., Hess, 2002). Governed by democratically elected officials, they were set up to respond to political signals rather than market signals. School choice, however, may actually undermine schools’ political responsiveness by promoting “exit” over “voice” for the most-active families. Because schooling is compulsory (unlike most market goods and services), there will always be a population of “non-choosers” whose fate is left to the conventional public schools (or to other schools of last resort, whatever they might be). Those students and their schools might be left substantially worse off if all of the most motivated and best informed families have exited to schools of choice.

Questions about the potential benefits of market forces and their potential harms cannot be resolved theoretically. Empirical evidence is needed to determine the extent to which competition can increase school productivity, on the one hand, and the extent to which it may produce stratification, sectarianism, and inequity, on the other hand. Choice and competition in education challenge the “common school” ideal that has been the model for American public education since the time of Horace Mann (Gill et al., 2007). This ideal model assumes that the public purposes of education demand that students in a community be educated under a common roof with a curriculum that promotes a common set of values. Departure from the common school ideal does not mean that the values associated with it are being abandoned, but it raises a critical empirical question about whether market competition can serve those values as well as (or better than) publicly operated schools can.

**Competition and Student Achievement**

We begin by examining the evidence of the effects of educational competition on public schools’ productivity as measured by students’ test scores.

**Competitive Effects of Charter Schools**

Charter schools have been in operation in the United States since the early 1990s, and the empirical evidence on their competitive effects on student achievement in conventional public schools is substantial. Table 13.1
summarizes studies that use student-level data sets to assess the competitive effects of charter schools. We exclude a few studies that have attempted to measure competitive effects using schoolwide data, because the growth of charter schools could change the composition of students in conventional public schools nearby, potentially biasing the results.

In the last decade, at least 10 studies have gathered longitudinal, student-level data from states and/or large school districts to attempt to measure effects of charter schools on the achievement of students in nearby conventional public schools. These studies have encompassed schools in at least 10 states. Their methods have varied somewhat, but they have in general examined changes in the achievement of individual students in conventional public schools that occur following increases in the number or market share of charter schools nearby. None of the studies have employed randomized experimental designs, which may be impossible to implement for purposes of measuring competitive effects (even though such designs can be quite valuable in measuring the direct effects of charter schools on their own students). Although the longitudinal, nonexperimental designs used are among the best available for real-world evaluations of competitive effects, they cannot provide definitive causal evidence because it is always possible that an unmeasured factor is actually driving the results.

Table 13.1 Studies of the Competitive Effects of Charter Schools

Among the existing studies of charter schools’ effects on students in nearby conventional public schools, most find either small positive effects or no effects. Only a single study (Imberman, 2011) in a single school district has found evidence of negative competitive effects of charter schools, and that study found some evidence of positive effects on student behavior even while it found negative effects in two of three academic subjects examined.

Several different explanations could potentially account for the positive competitive effect observed in several of these studies. The first possibility is that competitive pressure induces conventional public schools to increase their educational productivity. Alternately, competitive pressure may simply encourage a narrower focus on raising test scores. A third possibility relates to peer effects. If charter schools attract students with
lower-than-average achievement levels, the outflow of students to charter schools would not only produce an automatic increase in average achievement results in the conventional public schools, but could also produce gains for individual students who benefit from an increase in the average achievement levels of their peers. Still another possibility is that conventional public schools may experience an increase in per-pupil resources, if the resources lost with the departure of the charter students are less than the district’s average expenditures.

Even though the studies with findings of positive competitive effects outnumber those with findings of negative effects, the issue is certainly not settled. In fact the effects may vary across different communities, depending on the design of the charter law, the design of the local school funding regime, local demographics, and other factors. In one respect, the findings across all of the studies are consistent: no study has found large effects, positive or negative, of charters on conventional public schools. Future studies might seek to understand the importance of policy and financing to producing positive or negative competitive effects.

**Competitive Effects of Voucher Programs**

Table 13.2 summarizes the results of studies of the competitive effects of voucher programs in Milwaukee and Florida and a large-scale, quarter-century-old voucher program in Chile. As with the studies of charter schools, we include only studies that rely on student-level (rather than schoolwide) data.

The Milwaukee voucher program, enacted in 1990, provides students from low-income families with vouchers that can be used to attend private schools. For each student using a voucher, the school district loses approximately 30 percent of the state aid associated with that pupil. In 1998 the size of the program expanded dramatically, when an enrollment cap was raised from 1.5 percent of the total student population to 15 percent, and religious schools were allowed to participate, substantially increasing the available supply of spaces. It now enrolls over 15,000 voucher students in over 100 private schools.

Mader (2008) used longitudinal student-level data to examine the Milwaukee voucher program’s competitive effects, for math and reading scores in grades three through eight. He found that attending a school facing greater voucher threat was associated with higher math and reading scores, controlling for students’ prior test scores. The size of the effect varied by student type, with the smallest effect for low-income students.

Florida’s Opportunity Scholarship Program, launched in 1999–2000, was created as an integral part of the state’s high-stakes testing regime. It made students in the lowest-rated

---

Table 13.2 Studies of the Competitive Effects of Vouchers
Florida public schools (those that received "F" grades two years in a row) eligible for vouchers. During the first couple of years of the Florida voucher program, most of the schools that received an F grade improved in the next year, and thus avoided the voucher threat. In 1999–2000, 76 schools received their first F grade, but all of them improved the following year and received at least a D grade in 2000–01. In later years some schools received consecutive F grades and their students became eligible for vouchers.

West and Peterson (2005), Figlio and Rouse (2006), and Rouse et al. (2013) all found evidence that scores on Florida's state accountability test increased in public schools subject to the voucher threat. But two of the studies (West & Peterson, 2005; Figlio & Rouse, 2006) found no evidence of a similar effect on low-stakes test results in the same schools, leaving some doubt about whether the effects are related to narrow, test-taking skills. Moreover, even if the gains on the high-stakes test represent genuine learning, it is unclear whether the effects can be attributed to the voucher threat per se. Given that the voucher threat is an integral part of the high-stakes testing program, there is ultimately no way to distinguish the effect of the voucher component from the effect of the complete accountability system. The safe conclusion is that the high-stakes testing system as a whole has induced rising high-stakes test scores in targeted schools, and that the voucher threat may contribute to the incentives created by the high-stakes regime. Indeed, the fact that the strongest evidence of a positive competitive effect of vouchers comes from Florida is probably related to the design of the program, which gives public schools a clear incentive to respond by raising test scores (unlike in Milwaukee).

Winters and Greene (2011) examined effects of voucher threat from Florida's McKay Scholarship Program, which provides private-school scholarships to students with disabilities, on student achievement at nearby public schools. The McKay Scholarship Program went statewide in Florida in 2000–01, and was the nation’s first voucher program for disabled students. Using student-level data for 2001–02 through 2004–05, they found that competition from the voucher program was positively related to academic achievement in the public schools.

Figlio and Hart (2014) looked at effects of voucher threat from the Florida Tax Credit Scholarship program, which began in 2001–02. Using a student-level panel through 2006–07, they found positive effects of voucher threat on student math and reading scores, controlling for prior performance.

Voucher programs have been implemented outside the United States. With respect to the competitive effects
of vouchers, Chile has received more scholarly attention than other countries. Beginning in 1981, Chile gave
every student the option to use a voucher to attend a private school and made public school funding directly
related to the school’s enrollment. In response, more than a thousand new private schools opened in Chile, and
private school enrollment increased from 20 percent to 40 percent.

McEwan and Carnoy (2000) used a national student-level panel dataset to estimate the impact on public
school achievement of having competition from private schools serving voucher students. They found some
evidence of a positive effect of voucher competition on math and reading scores in the Santiago metropolitan
area, but slightly negative effects in the rest of the country. They speculated that the positive effects in
Santiago could be due to higher population density there, making private school competition more effective.
Their results are consistent with finding no aggregate effect for the country as a whole.

During the 1990s, the government of Sweden instituted a variety of education reforms nationwide, aiming to
decentralize educational authority, permit parental choice in schooling, and promote competition among
schools. Among other changes, the reforms created new subsidies for students to attend private schools; in
metropolitan areas, the proportion of Swedes enrolled in private schools increased substantially (Bjorklünd et
al., 2005). In a 2005 volume, Anders Bjorklünd and colleagues report that decentralization, choice, and
competition produced small positive achievement effects for most students in Sweden. They also sound a note
of caution, however: disparities in academic achievement in Sweden also increased—modestly—at the same
time.

In sum, evidence on the competitive effects of vouchers remains limited, but the existing evidence provides
reason for cautious optimism. No studies have found substantial negative effects, and a few have provided
evidence of positive effects. Evidence from Chile and Sweden suggests the importance of examining
differential impacts for different groups of schools and students, as well as average impacts. Finally, the
specific design of voucher programs is likely to be important in determining whether they have competitive
effects.

Tiebout Competition and Student Achievement

One of the most widespread potential sources of school competition is competition among public schools
themselves. One source of this public school competition would be “Tiebout choice,” describing residential
choices made by parents, which may be based partly on the quality of the public schools in that neighborhood.
This could induce indirect competition among school districts, as they compete for additional local property
tax revenues and per-student funding. There are few empirical studies of the effect of Tiebout choice, however.
Examining the effect of this kind of naturally occurring choice is very difficult, because it is impossible to
conduct a randomized experiment or even to observe changes in student achievement before and after an
intervention occurs. In consequence, causal inferences are problematic.

Among the few studies of the effects of Tiebout choice, perhaps the most well known is Hoxby (2000), which
assessed the impact of public school competition on student achievement by examining nationwide variation in
the amount of public school choice available in different metropolitan areas. Because local school district
structure is related to other factors that influence student achievement, Hoxby used the number of rivers and
streams in the metropolitan area as an instrument, in order to determine the portion of variation in public
school choice that is independent of other factors affecting student performance. She found that having more
public school choice is related to higher levels of student performance, controlling for other local
characteristics. Rothstein (2005), however, found that Hoxby’s results were not robust across different
specifications.

Using data for elementary schools in the San Francisco Bay area, Bayer and McMillan (2005) estimated the
change in demand each school faces in response to a change in that school’s quality. They related this measure
of local competitiveness to student achievement, controlling for observable school characteristics, and found
that a one standard deviation increase in the competitiveness of a school’s local market was associated with a
0.15 standard deviation increase in school performance.

Card and Rothstein (2007) use student-level data on SAT scores across different regions to estimate how
much of the variation in student performance is due to school and neighborhood segregation, using cohorts
from 1998 through 2001. They find that the black-white test score gap is higher for more segregated cities, and
that much of the effect of neighborhood segregation is due to segregation by income, rather than race.

In sum, the competitive effect of Tiebout choice on student achievement is not clear; different studies have produced conflicting results. Given the methodological difficulty of analyzing the competitive effect of Tiebout choice, it is likely to remain a matter of debate.

**Competition, Integration, and Civic Socialization**

Although a growing number of studies has attempted to measure the effects of competition on achievement in math and reading, relatively little attention has been devoted to the effects of competition on two of the traditional public purposes of schooling: social integration and civic socialization. Because choice and competition represent fundamental departures from the common-school model for the provision of public education—a model intended (at least in its ideal) to promote integration and civic socialization—empirical evidence on these issues is especially important.

**Competition and Student Sorting**

A few scholars have used theoretical models to predict the effects of school-choice policies on the distribution of students. Epple and Romano (1998) predicted that flat-rate educational vouchers would increase the stratification of students by wealth and ability, in the absence of constraints on tuition and admissions policies in private schools. High-income, well-informed, motivated parents would find their way to private schools that would be seeking high-achieving, motivated students whose families could afford to "top off" the voucher with additional tuition payments. In a later paper, however, the same authors concluded that school-choice policies could be designed in ways to "reap the benefits of increased competition without increased stratification" (Epple & Romano, 2002, abstract). In particular, they predict that voucher policies could avoid increasing stratification of students by requiring participating private schools to accept the voucher as full payment of tuition (i.e., to prohibit "topping off" the voucher with additional tuition payment—a prohibition that is in fact incorporated in some existing voucher policies) and by varying the voucher amount to provide larger scholarships to students with greater educational needs (much as the state of Florida does in its McKay scholarship program for students with disabilities).

School-choice policies may have different effects on integration in neighborhoods than on integration in schools. Nechyba (1999, 2000, 2003) used theoretical models to predict that vouchers should reduce residential stratification, because they cut the link between school quality and residential location for families using them. Access to vouchers to attend private schools will make some families more willing to live in communities with low-performing public schools—communities that otherwise would be populated largely with low-income families. Ferreyra (2007) found similar results when simulating the effects of a high-value, universally available voucher for residents of Chicago.

Few empirical studies have systematically examined the effects of charter schools on the stratification of students. Some reports (e.g., Frankenberg, Siegel-Hawley, & Wang, 2011) have descriptively compared the racial/ethnic composition of charter schools and conventional public schools in their states. Such studies, however, cannot estimate the effect of charter schools on integration, because they provide no evidence on how students would be distributed in the absence of charter schools. Moreover, they sometimes compare enrollments in charter schools with district-wide or statewide averages, without recognizing that charter schools often open in neighborhoods that are highly segregated.

The few studies that have carefully examined stratification issues related to charter schools all have found some evidence that charter schools may increase stratification by race—but in total the studies suggest that stratification effects may vary by racial/ethnic group and by local context. Bifulco and Ladd (2006) found that charter schools in North Carolina have contributed to an increase in racial stratification, with both white and black students becoming more racially isolated. Booker et al. (2008) found that in California and Texas, black students transferring from conventional public schools to charter schools tended to move to schools with higher proportions of their own racial group and lower levels of racial/ethnic integration. Movements of white
and Hispanic students, by contrast, had mixed effects on schools’ integration. Ross (2005) examined statewide data in Michigan and found that the exposure of black and Hispanic students to white students in conventional public schools declined in districts that had large proportions of students enrolling in charter schools. Zimmer et al. (2009) looked at student movement to charter schools in Chicago, San Diego, Philadelphia, Denver, Milwaukee, Ohio, and Texas. They found that generally students transferring to charter schools did not lead to large changes in the sorting of students by race/ethnicity. There was some variation, with charter schools marginally reducing racial integration in Philadelphia and Texas, and marginally increasing racial integration in Chicago. Also, they found some evidence that black students are more likely to move to charter schools with higher concentrations of black students than the schools they left, in five of the seven sites.

Scholars have not used student-level data to examine the effects on racial and ethnic distributions of movements of students from public to private schools under the voucher programs that have been operating in Milwaukee, Cleveland, Florida, and Washington, DC. In three of these cities, however, studies have compared existing levels of racial stratification in voucher and charter schools. In Milwaukee (Fuller & Mitchell, 1999, 2000; Fuller & Greiveldinger, 2002) and DC (Greene & Winters, 2007), the studies found evidence that students participating in the voucher programs were somewhat less likely to be in schools that are highly stratified racially than were students in public schools. In Cleveland, Greene (1999) found a mixed picture: voucher students were slightly less likely to attend schools enrolling over 90 percent minority students, but they were more likely to attend schools enrolling over 90 percent white students. As compared with public school students in the city of Cleveland and in the surrounding suburbs, Cleveland voucher students were more likely to attend schools with racial distributions near the region-wide average.

Other evidence of the effects of school-choice policies on the stratification of students is available from policies in place outside the United States. New Zealand created a comprehensive system of universal choice among publicly supported schools in 1991, while giving the schools substantial autonomy in operations—thereby creating a competitive market in education. The system allows oversubscribed schools to choose students. Ladd and Fiske (2001) found that the resulting distribution of students across schools in New Zealand was highly stratified by ethnicity, with Maori and Pacific Islander students disproportionately concentrated in schools that are not in high demand.

Willms (1996) examined sorting effects over a decade after the introduction of school choice in Scotland in 1991. Unlike the system in New Zealand, Scotland’s system is not a universal choice system, but one that permits choices by families seeking them out. Willms found that choosing parents tended to move their children to schools with higher achievement levels and higher socioeconomic status than the assigned schools they left. In consequence, communities with high proportions of students making active choices saw the largest increases in stratification by socioeconomic status in schools.

In sum, studies of the effects of school-choice policies on the sorting of students suggest that stratification across schools may increase under some circumstances. It also appears likely, however, that the stratification effects of school choice programs differ depending on the local context and the policy design. In particular, universally available school choice programs may reduce integration across racial/ethnic or socioeconomic lines, but programs targeted to low-income families could conceivably increase integration, particularly if implemented in communities where public schools are presently highly stratified.

Nearly all of the empirical studies of the effect of school competition on student stratification take as their starting point the assumption that the school is the appropriate unit of analysis. But student sorting can occur within schools as well as between schools. If integration is to benefit students, presumably it requires that they interact in the classroom, the cafeteria, and the gym, rather than having separate educational experiences that happen to occur under the same roof. In many large public high schools, social factors and the sorting of students into discrete academic “tracks” combine to make students’ daily interactions less integrated than schoolwide counts of students might suggest. Unfortunately, no studies have compared the internal integration of charter schools, voucher schools, and conventional public schools. One study (Greene & Mellow, 1998) compared integration in cafeterias in samples of public and private schools in two unnamed cities and found that private schools were somewhat more likely than public schools to have students of differing racial/ethnic groups sitting in close proximity. Whether this finding would generalize to other cities, and whether it predicts what would occur under policies promoting school choice, are unclear. More research on integration inside schools is much needed.
Competition and Civic Socialization

Integration has typically been regarded as a public good under the assumption that students who are exposed to others of different backgrounds will develop a sense of tolerance and a wider community allegiance. This assumption is implicit in the common-school model: the education of students in common is presumed to foster healthy civic attitudes. Moreover, the common-school model also expects that healthy civic attitudes will be promoted because public schools are ultimately accountable to democratically elected officials, who will ensure that the schools have a curriculum that explicitly promotes civic values (see McDonnell, Timpane, & Benjamin, 2000). Nonetheless, some charter schools and charter management organizations (such as Democracy Prep and the Cesar Chavez Schools for Public Policy) have explicitly made the education of citizens their primary purpose (Lake & Miller, 2012).

Despite these long-standing assumptions about the benefits of the common-school model, there is little empirical evidence on the point. Meanwhile, few scholars studying vouchers and charter schools have examined civic outcomes (Gill et al., 2007). The debate about school choice and civic values has been conducted almost entirely in the legal arena, where advocates on both sides have argued whether permitting public funding to go to religiously affiliated private schools violates provisions of state and federal constitutions. The constitutional debate makes no reference to empirical evidence on the civic outcomes produced by publicly and privately operated schools, and it obscures the possibility that the issue may extend beyond religion, affecting charter schools as well as private schools.

Only one study has compared charter schools and conventional public schools in terms of effects on students’ attainment of civic skills and knowledge. Buckley and Schneider (2004) used results from telephone surveys of students in grades seven through 12 to compare the values, civic participation, and knowledge of charter students and students in conventional public schools in the District of Columbia. They found that charter students were significantly more likely than students in conventional public schools to perform community service, while they had no greater probability of participating in clubs or sports. Charter students were marginally significantly more likely to take part in a debate or discussion and to speak at a community meeting. They found no differences between charter students and students in conventional public schools in terms of political tolerance. Buckley and Schneider’s findings regarding the potential advantages of charter schools for civic skills and community service are provocative, but given the high rate of attrition from the sample and the difficulty of controlling for selection bias, they should be viewed as tentative.

No studies have yet examined the effects of voucher programs on civic outcomes, but three studies have assessed the effects of private schools. Campbell (2001) and Belfield (2003) used data from two different waves of the National Household Education Survey (NHES) to examine the effects of public and private schools on students’ civic skills, civic confidence, political knowledge, political tolerance, and participation in community service. Using a cross-sectional regression analysis that controlled for a wide range of student, family, and school characteristics with 1996 NHES data, Campbell found that Catholic schools (but not other private schools) had a positive effect on community service. Catholic schools were also associated with greater civic skills and civic knowledge. Catholic schools, other religious schools, and secular private schools were all found to have advantages over public schools for civic confidence (i.e., confidence in one’s ability to write a letter to a government official or to make a statement at a public meeting). Catholic schools and secular private schools had students with greater political tolerance than did public schools, while students in non-Catholic religious private schools had less political tolerance. Belfield (2003) found similar results using similar methods and data from a subsequent administration of the NHES in 1999.

Godwin, Godwin, and Martinez-Ebers (2004) surveyed high school students in public schools and fundamentalist Christian schools in one metropolitan area, and compared the political values and attitudes of students in both types of schools in 10th grade and 12th grade. Controlling for various family characteristics, 10th-grade students in fundamentalist Christian schools had lower scores than did public-school students in terms of some measures of tolerance. Because the 12th-grade cohort of students in fundamentalist schools exhibited substantially higher levels of tolerance, however, the difference in tolerance between grades 10 and 12 in fundamentalist schools exceeded that in the public schools. The authors interpret this quasi-longitudinal result to suggest that “Fundamentalist schools appear to be as successful as public schools in teaching the values necessary to assume the burdens of citizenship in a democratic society” (p. 1109).

201
In the absence of randomized designs and true longitudinal data, all three of the studies on civic socialization in private schools should be viewed as providing only suggestive rather than conclusive evidence. Nonetheless, the studies so far provide little evidence to confirm fears that privately operated schools might promote attitudes unsuitable to socialization in a democracy. In some instances, they suggest reason for optimism that some kinds of privately operated schools could actually improve public outcomes.²

Such optimism also gains support from the experience of European countries that have long experience with public funding of privately operated religious schools. In some countries—notably the Netherlands—providing public support for the religious education preferences of parents has been regarded not as a threat to civic unity but as an essential support for it. A recent collection of essays edited by Wolf and Macedo (2004) describes the history, politics, and policy behind the varied systems of public support for private schools that exist in the Netherlands, Belgium, Germany, France, the United Kingdom, and Alberta, Canada—as well as some empirical evidence on the effects of the private schools on civic values. The chapter by Dronkers (2004) reviews empirical evidence from studies of private schools across continental Europe, finding no evidence that the religious schools promote different civic values than do the government-operated schools. As all of the case study essays in the volume make clear, however, the “private” religious schools that receive public funding in European countries are subject to far more regulation than are private schools in the United States. In Europe, public funding has come with extensive public regulation. Whether this kind of regulation is necessary to ensure the promotion of civic outcomes in private schools is unclear. And whether it would be constitutionally permissible and politically acceptable in the United States is also unclear. Nonetheless, at minimum the other countries provide useful examples for considering how U.S. policymakers might promote civic purposes in privately managed and religious schools.

Policy Implications and Future Research

Competitive effects—positive or negative—are among the most important outcomes that may be produced by educational privatization and school choice. Effects on students remaining in conventional public schools could easily dwarf direct effects on students who make use of school-choice policies. Moreover, privatization of the governance of schools may have implications for the public purposes of schooling that are not often considered in research on school-choice programs. A few studies, however, have provided some guidance on these issues.

Existing research on charter schools and voucher programs suggests reason for cautious optimism about their competitive effects on the achievement of students in conventional public schools: the studies finding positive effects or no harm outnumber those finding negative effects. The literature provides more reason for concern about the effect of school choice on the stratification of students across schools. Some varieties of choice programs appear to increase stratification by race or socioeconomic status.

Regrettably little empirical evidence is available on the effect of competition and privatization on one of the long-recognized public purposes of education: the socialization of citizens. There is as yet no empirical evidence that supports the hypothesis that serving this public purpose requires schools that are publicly governed. The very limited research base suggests the intriguing possibility that some kinds of privately operated schools may promote tolerant attitudes, civic knowledge, and community service more effectively, on average, than do conventional public schools. But the evidence on this point is thin indeed, and researchers have not had the opportunity to apply the kinds of longitudinal methods that have been used to assess impacts on math and reading achievement.

Indeed, more research on all of these points would help to clarify the effects of competition on a variety of outcomes and under a variety of circumstances. Effects on outcomes other than math and reading achievement are especially in need of more study. Competitive, decentralized models of K–12 education delivery such as those represented by charter schools and vouchers involve a fundamental departure from the standard, “common school” model that has dominated American schooling for the past two centuries, so the relative absence of attention to the public purposes that the common-school model aims to achieve is especially notable.

Additional research should also more deeply explore the measurement of competition and competitive effects. Future work could include closer attention to patterns of integration inside schools as well as between
them. In addition, the measure of competition itself merits examination. Levacic (2004) points out that structural measures of competition, such as the percentage of students in an area attending alternative schools or the number of alternative schools within a specified geographic area, are not necessarily related to the degree of competitive pressure actually felt by the decision makers at the public school or district. Examining schools in England, she finds that the degree of competition perceived by school officials has a strong positive effect on student performance but that structural measures of competition do not.

Despite the limitations of the existing evidence on the effects of competition, enough exists to provide some guidance to policymakers. The specific design of a program of educational competition is likely to matter. The effect of school choice on integration will depend substantially on what students and schools are eligible to participate in the program. Some choice programs that are open to all students have had the effect of increasing the stratification of students across schools. Voucher programs that are targeted to low-income students, by contrast, have the potential to reduce stratification. That outcome will occur if they succeed in bringing low-income and minority students into private schools that presently enroll higher-income, tuition-paying students.

Other predictions about the effects of policy design may not follow directly from the existing empirical literature, but can be logically inferred with plausible assumptions. For example, universally available voucher programs that provide small scholarships (i.e., less than is necessary to pay full tuition at many private schools) could easily exacerbate stratification by subsidizing upper-income families who can afford to pay tuition above the scholarship amount. And choice programs that allow participating schools to select their students might have more-negative effects on stratification than would programs that require participating schools to accept all applicants (or to choose randomly among them if they are oversubscribed). Finally, choice programs may have more success in inducing improved achievement in conventional public schools if they are explicitly designed with that aim in mind, by incorporating accountability systems (as does Florida’s A+ Schools program), by removing some constraints on public schools that might reduce their ability to compete, or by increasing financial incentives for improvement by ensuring that educational funds follow each student to whatever school he or she attends.

Less can be said with confidence about the specific policy levers that might promote favorable effects on civic socialization. Too little is known about the effects of different kinds of schools on this outcome. At minimum, it would make sense for choice programs to attempt to measure civic socialization in participating schools and to make public any information about the civic components of schools’ missions and curricula.

Indeed, provision of good information about schools—going beyond achievement results in reading and mathematics—should be viewed as an essential component of any educational program based on choice and competition. High-quality information for consumers (i.e., parents and students) is a critical ingredient for the effective operation of markets. Better information therefore merits serious attention from policymakers interested in educational competition—not only because additional research is needed on the effectiveness of choice programs and schools, but also because information is itself a fundamental component of a competition-based system.

Notes

1. On the disproportionate representation of racial/ethnic minority students in lower academic tracks, see Brad-dock (1990); Braddock and Dawkins (1993); Gamoran (1987); Oakes (1985, 1990).

2. If privately operated schools had the effect of increasing educational attainment, they might have an indirect and positive effect on civic outcomes as well. Dee (2004) has found that greater educational attainment may increase voter participation, civic knowledge, and support for free speech. Whether charter schools or voucher schools increase the educational attainment of their students has not yet been empirically examined.

References


Fuller, H. L., & Greiveldinger, D. (2002). The impact of school choice on racial integration in Milwaukee private


New York: National Center for the Study of Privatization in Education.


Section III
Promoting Equity and Adequacy

SECTION EDITOR: LEANNA STIEFEL
Introduction

This chapter provides an overview of recent literature on conceptions of educational equity and adequacy applied to state school finance policy in the United States. Equity conceptions deal primarily with variations or relative differences in educational resources, processes, and outcomes across children, whereas adequacy conceptions attempt to address in more absolute terms how much funding, how many resources and what quality of educational outcomes are sufficient to meet state constitutional mandates. State constitutions address explicitly state legislatures’ responsibility toward public schooling. The fact that the U.S. Constitution does not do so, however, leads to significant limitations for resolving inequities or inadequacies in public schooling nationwide. Equity and adequacy remain issues governed primarily by state courts and legislatures.

The chapter draws on work by scholars in the areas of public finance, law, and school finance policy. We begin the second section with a discussion of Robert Berne and Leanna Stiefel’s framework for evaluating school finance policy introduced in 1984 following initial waves of legal challenges to state school finance policies. We then trace the roots of the conceptions laid out by Berne and Stiefel, including the public finance policy roots of horizontal and vertical equity and early legal theories of equal educational opportunity. The third section of this chapter addresses vertical equity and the intersection between conceptions of vertical equity and educational adequacy. We point out that vertical equity and adequacy conceptions are separable but have largely been folded into one as a matter of legal convenience in the context of state education clause challenges to school finance formulas. The fourth section provides a synthesis of equity and adequacy concepts in relation to law. The final section addresses issues for future consideration.

Origins of Equity Conceptions in School Finance

School finance policy lies at the intersection of the fields of public finance, law, and other social sciences. As first comprehensively framed by Berne and Stiefel (1984), school finance equity may be viewed from either or both the perspectives of the taxpayer and/or the child. Because equity measurement evaluates the distribution of objects, there exists the critical underlying question—across which units of organization should the distribution of objects be evaluated? Ideally, equity would be measured across children or taxpayers using precise measures of the educational inputs available to each child or tax burden shouldered by each individual taxpayer. Most U.S. states organize public schooling into local education agencies, or school districts with school buildings within districts. Hawaii, which operates as a single statewide district, is the notable exception. In most U.S. states, the primary role for the intermediate governance unit—the school district—is financial management, including authority over local taxation for annual operations of schools and infrastructure. State government interventions to resolve inequities have been designed primarily to compensate inequities in revenue raising capacity across school districts (see Picus, Goertz, & Odden, this volume, for more detail). State aid is allocated primarily to districts rather than to schools or children. As such, most school finance distributional analyses focus on distribution across school districts, masking potentially large disparities across school buildings and children within buildings.

Berne and Stiefel (1984) offer the following four guiding questions for evaluating equity in school finance:
Who? What? How? How much? Under the question of *Who?*, various constituents might be addressed, including children, teachers, and taxpayers. Most often, children, clustered within school districts, have been the emphasis in school finance equity analyses.

Under the question of *What?*, one may focus on (1) financial inputs to schooling; (2) resources purchased with financial inputs that include teachers, equipment, materials, and supplies, as well as facilities and transportation; (3) educational processes, including time spent on specific activities, or student participation rates in specific courses of study; or (4) student outcomes, ranging from measures of academic achievement to measures of economic productivity. In practice, analyses have focused on the first of these: financial inputs to schooling, usually per-pupil expenditure, measured at the school district (local education agency) level. Further, most analysts have evaluated state-level systems of financing public education, excluding children in private and home schooling. Cross-state analyses have been relatively rare, as have analyses that include private schooling or more recently charter schooling.

In response to the question of *How?*, Berne and Stiefel adapted public finance concepts of horizontal equity defined as the *equal treatment of equals*, and vertical equity defined as the *unequal treatment of unequals*, and added to these concepts *equal educational opportunity*. Berne and Stiefel defined equal opportunity to include what is now typically called fiscal neutrality. Fiscal neutrality means that variations in resources across children should not be a function of the wealth of the community in which a child happens to live. Horizontal equity states that resources should be equally available to all students attending school within a state, provided all students have equal needs. Vertical equity applies to those cases where specific students or groups of students have identifiably different educational needs and where meeting those needs requires additional resources.

The *How much?* question requires statistical measures, which can be distinguished by their emphasis on different ranges of the resource distribution. These measures address questions such as how much variation in resources across similarly situated students is acceptable? How much variation across differently situated students is necessary and on what basis?

**Historical Origins of the Berne and Stiefel Framework**

Theories underlying Berne and Stiefel’s framework came from two discrete sources that were adapted for evaluating the raising of public education revenue and distribution of expenditures across public schools. The first of these sources was public finance principles applied to tax policy. The second was legal theory. Given this context, the frameworks addressed herein are unique to the public education system in the United States, a nation where defining educational rights has been left largely to the states and where no broader guarantee of education as a basic human right is explicitly acknowledged (see Fiske & Ladd, this volume, for contrast with developing countries).

**Public Finance Origins**

Though applied by Berne and Stiefel to the expenditure side of the school finance equation, the basic concepts of horizontal and vertical equity were drawn from a lengthy literature in public finance that focuses on tax policy. Within the context of this literature, horizontal equity refers to the equal treatment of equals and is a concept around which there is little dispute. Simons (1950, p. 8) notes “it is generally agreed that taxes should bear similarly upon all people in similar circumstances.”

Like its counterpart in school finance, the concept of vertical equity in tax policy is more controversial. In general, vertical equity (VE) in tax policy refers to the progressivity of a tax, or how much more certain individuals should be asked to pay in taxes, and on what basis. The well-known public finance economist Richard Musgrave (1990, p. 113) notes:

VE (Vertical Equity) on the contrary is inherently controversial. An appropriate pattern of differentiation must be chosen but people will disagree on its shape. Whereas HE (Horizontal Equity) is a minimal rule of fairness, VE (Vertical Equity) is a matter of social taste and political debate.
Further, some argue that there is little or no need for a separate conception of vertical equity, if horizontal equity applies specifically to equal treatment only of equals. As such, horizontal equity accepts that unequals should be treated unequally. Musgrave (1959) suggests that “the requirements of horizontal and vertical equity are but different sides of the same coin” (p. 161).

Scholars of tax policy suggest two possible standards for determining the fair treatment of taxpayers that are unequal: (a) the benefit standard, and (b) the ability standard. Under the benefit standard, individuals or groups who stand to reap the greatest benefit from a tax should pay higher taxes. Under the ability standard, individuals with greater ability to pay a tax should pay higher taxes relative to their income, as would be the case with a progressively structured income tax. In the tax policy context, one is still left with the value-laden concern over just how progressive is progressive enough? We suggest at a later point that this vertical equity question is more easily answered in the context of modern public education finance.

Legal Origins

Providing additional backdrop to Berne and Stiefel’s framework was the emergence of early legal theories on how best to challenge and resolve significant disparities across school districts, schools, or groups of students in the quality of public schooling. Specifically, in the late 1960s legal theorists explored how one might challenge disparities in the quality of schooling across children, disparities that persisted through the period of racial integration of public schooling in the first few decades following Brown v. Board of Education. A flurry of activity emerged in the late 1960s following implementation of the Civil Rights Act of 1964, the Elementary and Secondary Education Act (1965), and release of the Coleman Report, Equality of Educational Opportunity (1966).

Even in their early stages, challenges to school finance formulas required evaluating the nature as well as the existence of disparities. Were differences in resource distribution related to legitimate vertical equity concerns? Were they simply arbitrary? Or were these disparities associated with some factor courts could identify as offensive, such as race or ethnicity? The political wounds of Brown v. Board still being fresh and the remedy response slow if not entirely stalled in some states, scholars sought ways to define disparities across groups without reference to race (Ryan, 1999).

Thus, legal theorists of the late 1960s suggested frameworks for evaluating variation in educational resources, with the particular goal of showing that resource variation deprived classes of students (including groups classified by race) of access to their fundamental right to education. Unfortunately, because such a fundamental right was not explicitly spelled out in the U.S. Constitution, the possibility of federal court challenges to the overall adequacy of educational funding and the scope of subsequent equal protection challenges was severely limited. Early frameworks sought to build on the Supreme Court’s treatment of the Equal Protection Clause in the 1950s and 1960s (Enrich, 1995). Especially important were the Court’s emphasis on the importance of education in American life (Brown v. Board), particularly the Court’s suggestion that governments had an affirmative duty to ensure that governmental services were equally available to all.

Horowitz (1966) suggested that courts consider variations in resources that were geographically arbitrary to violate equal protection—the principle of Geographic Uniformity. Coons, Clune, and Sugarman (1969), building on this argument, proposed the principle of fiscal neutrality (discussed in Berne & Stiefel, 1984). They argued that children in property-poor school districts were discriminated against by state policies that allowed heavy reliance on local property taxation for public schooling.

The Coons, Clune, and Sugarman framework provided the basis for arguments in two seminal early school finance cases: Serrano v. Priest in California state court and San Antonio Independent School District v. Rodriguez in federal court. Both were equal protection challenges, and both challenged disparities in funding across school districts that resulted from differences in local taxable property wealth and local taxing decisions under state policy. The California State Supreme Court agreed that education was a fundamental right and, further, that property-wealth related disparities in school district funding were unacceptable, in part on the assumption that wealth was a suspect class and that differential treatment by wealth should be reviewed under strict scrutiny. The U.S. Supreme Court viewed wealth-related school funding disparities in Texas differently. It ruled that under the U.S. Constitution, education was not a fundamental right, because it is not explicitly
mentioned in the constitution, and further that individuals residing in property-poor school districts were not a suspect class; therefore only rational basis scrutiny was required. Ultimately, the U.S. Supreme Court accepted that the State of Texas had a legitimate interest in maintaining locally controlled taxation for public schooling and chose not to overturn the state's school finance policy.

Arguably, Rodriguez was not the best case for early federal court application of the Coons, Clune, and Sugarman framework because it could be too easily shown, in Texas in particular, that taxable property wealth per pupil at the school district level was not highly associated with financial well-being of individual families and children residing in those districts. Testimony in Rodriguez revealed the complexities of evaluating the relationship between school funding and community attributes. Unfortunately, while these complexities were somewhat context specific, the court's ruling in the case was broad, shutting off all future federal court challenges to property-wealth related disparities in educational funding. While the arguments failed in federal court, there were some additional early successes in state courts beyond Serrano.10

The Shift from Input Equity to Outcome Equity in School Finance

The shift toward greater emphasis on vertical equity is in large part associated with a shift away from a focus on the equality of educational inputs toward the equality of educational outcomes. The publication of James Coleman’s report, Equality of Educational Opportunity, in 1966 is often cited as a major turning point. Among other things, Coleman and colleagues highlighted the strength of the influence of family background characteristics on student outcomes and the relative insensitivity of student outcomes to school and district resources.

While shifting focus toward educational outcomes, the Coleman report raised questions that would linger through current school finance litigation about the tenuous nature of the relationship between schooling quality—as measured by student outcomes—and financial inputs to schooling. Does money matter? Will equal inputs lead to equity defined in terms of educational outcomes? Or are differential inputs needed to account for different needs? Early state court cases overturning funding disparities accepted that expenditures on schooling were related to educational quality: "[we] accept the proposition that the quality of educational opportunity does depend in substantial measure upon the number of dollars invested."11 Later state court cases addressed more directly the assumptions that (1) money in general does matter for improving student outcomes and (2) money may matter more for some children than for others, which represents an endorsement of the concept of vertical equity.12

Vertical Equity as Outcome Equity: Setting the “How Much” Standard

Recall that in the tax policy literature no standard provides a decisive answer to the question of “how much” progressivity is enough. Arguably, in school finance policy, the "how much" question is easier to address. At least at the conceptual level, how much, in terms of the differential inputs required for groups with different educational needs, can be defined in terms of the relative sufficiency of those resources toward improving equity of educational outcomes. Those outcomes may be measured either in normative terms or against specific criteria. Such criteria might include acceptable minimum performance on state assessments, graduation rates, or successful matriculation to and persistence through postsecondary education. In that sense, vertical equity as a school finance conception might simply be redefined as horizontal equity of outcomes or horizontal equity of opportunity to achieve a defined outcome. Stated differently, equality of outcomes requires differentiation of inputs.

Shifting the focus toward outcomes, however, raises the question of which outcomes are of greatest interest. The most common outcome in school finance litigation is student achievement as measured by state standardized test scores. Test scores have become the centerpiece of school finance litigation in part because of their role in state performance-based accountability systems. Further, most statistical analyses of the costs of achieving state-mandated educational outcome levels have focused on the costs either of achieving specific test score levels on state assessments or on the costs of achieving specific percentages of children scoring proficient
or higher on state assessments. Less attention has been paid to longer term outcomes such as college attendance and completion, future earnings, or harder-to-measure quality of life outcomes addressed in disability literature (Turnbull, Turnbull, Wehmeyer, & Park, 2003). In contrast to the narrow focus on test scores in cost studies, judicial decrees based on state education clauses have expressed a far broader conception of the desired outcomes of schooling, focusing most often on economic and civic participation.

Despite the failure of equal outcome claims in federal courts under the equal protection clause, this outcome perspective remains the most conceptually, empirically, and legally viable approach for addressing the “how much” question of vertical equity. To accept an alternative view, such as requiring state and local governments to provide only reasonable accommodations in terms of educational inputs, where those inputs may knowingly be insufficient to promote equal or adequate outcomes, is to suggest that state governments should be granted wide latitude to establish different outcome standards for some children than for others.

The federal No Child Left Behind Act of 2001 (NCLB) provides further clarity on this point by requiring that states adopt outcome measurement systems that apply to all eligible students. Further, NCLB requires that states not only monitor achievement gaps across children by race, ethnicity, language proficiency, and poverty, but also that states close achievement gaps between these groups over time. Most states have now promulgated systems of outcome measurement and accountability compliant with NCLB, and some have gone beyond NCLB requirements to tie measures of individual students’ achievement to the granting of a high school diploma.

The allowance of numerous accommodations and exceptions under NCLB for children with disabilities along with the independent statutory framework (IDEA) for classified children means there are two separate systems for determining vertical equity for children with disabilities: one outcome-based (NCLB) and one due process and predominantly input-based (IDEA). In part this difference reflects the practical concern over the extreme cases in which a child’s disabilities are so severe that no level of resources would be sufficient to generate a high academic outcome. The conflict can only be resolved with significant resources targeted to those extreme cases, coupled with standards so low as to be meaningless for most children.

Educational Adequacy and Vertical Equity

Due to failures of the “equal protection” doctrine under the U.S. Constitution to guarantee vertical equity, arguments for vertical equity have found their recent legal home in state education clause challenges, which increasingly focus on the adequacy of funding provided by states to local public school districts. Typically, adequacy is measured in terms of whether funding is sufficient for children in schools to achieve state-mandated minimum outcome levels (based on systems promulgated under NCLB). These accountability systems vary across states. Where outcome equity is emphasized, differentiation of inputs across districts with different proportions of hard-to-educate children is required.

On the heels of William Clune’s influential article “The Shift from Equity to Adequacy in School Finance” (1994), Julie Underwood (1995) framed vertical equity as synonymous with educational adequacy. Underwood asserts that each individual child requires a specific set of educational programs and services to achieve the desired educational outcomes for that child. As such, those inputs are necessarily differentiated. In fact, under this conception, while horizontal equity theoretically exists, it may not exist in practice, since no two children’s educational needs are exactly the same.

Underwood’s framework, while useful, applies under very limited conditions. It applies only to vertical equity pertaining to individual student educational needs, and it requires that a level of desired educational outcome be specified for each child or all children. Underwood’s framework is primarily an adequacy framework, pegged to a specific level of outcome, and with emphasis on state mandated outcome levels.

Beyond individual students’ needs, conditions that influence vertical equity apply to concentrations of children from economically disadvantaged backgrounds, children with limited English language proficiency, minority student concentrations, and average peer group prior achievement level. While Underwood’s conception addresses the student’s individual needs, her approach fails to acknowledge the potential effects of peers on students’ individual outcomes and of different costs of improving those outcomes as peer composition varies. Underwood’s conception also fails to address the variation in outcomes due to labor market costs (especially the cost of attaining sufficient quantity of teachers of specific quality) and economies of scale.
(whereby some districts have high costs due to small enrollments), among other things. Individual student background attributes are but one small piece of a complex integrated puzzle in which the specific educational needs of individual students interact with the composition of students’ peer groups and with the context in which children are schooled. These factors affect comprehensively the costs of achieving specific educational outcomes.

For any specific level of outcome, inputs must be distributed in a vertically equitable way to achieve that outcome. However, vertical equity remains separable from adequacy in that vertical equity is a purely relative concept. Vertical equity of inputs is necessary for providing either equal opportunity to achieve a specific outcome, or equal outcomes.\textsuperscript{11}

Koski and Reich (2006) offer an alternative rationale for a purer conception of vertical equity. Koski and Reich argue that the quality of educational outcomes is largely relative because education is primarily a competitive\textit{ positional good}. Hirsch (1976, p. 52) describes positional competition as follows:

By positional competition is meant competition that is fundamentally for a higher place within some explicit or implicit hierarchy and that thereby yields gains for some only by dint of loss for others. Positional competition, in the language of game theory, is a zero-sum game: what winners win, losers lose.

If education is considered to be a competitive positional good, the extent of disparity in educational outcomes above a minimum outcome standard codified in state policies matters a great deal. Baker (2005) describes variation in educational opportunity above the minimum standard as\textit{ opportunity surplus}. Under a \textit{vertical equity as adequacy} conception, it matters only whether individuals have sufficient resources to achieve a state mandated minimum outcome level. That is, only opportunity deficits must be erased. Opportunity surpluses are considered nonoffensive.

Koski and Reich argue, however, that the value of achieving that minimum standard is largely contingent on the variation of education opportunities above it. In a system where children are guaranteed only minimally adequate K–12 education, but where many receive far superior opportunities, those with only minimally adequate education will have limited opportunities in higher education or the workplace. Based on the concerns expressed by Wise (1976), it is likely that the least fair state school finance systems in the nation might surpass the minimum adequacy standard by allowing no one school or child to fall below a meaningless minimum outcome threshold, while allowing dramatic degrees of opportunity surplus beyond that minimum threshold.

As with a \textit{reasonable accommodations} view of vertical equity, the minimum adequacy compromise constitutes a state endorsement of different outcome standards leading to different life opportunities for different children. While some compromise may be an unavoidable reality of school finance policy, the compromise need not be embedded into the framework for evaluating fairness in state school finance policies. Evaluation frameworks are better built on pure equity conceptions, with compromises left for political deliberation over school finance and judicial evaluation of constitutional compliance.

**Synthesis and Summary**

We conclude this chapter with a framework in which we synthesize theoretical developments in evaluating the fairness and equity in state school finance formulas. We suggest that the evaluation of the equity and adequacy of modern state school finance formulas might be best addressed through the following four questions, approached sequentially:

1. Is there variation in educational inputs such as per-pupil expenditures across districts and children?
2. What is the nature of that variation? From a legal standpoint, it remains relevant to discern whether variation in resources across school districts is (a) related to fiscal capacity and local control issues, (b) related to vertical equity or marginal cost issues, or (c) otherwise unjustifiable variation created by legislative actions. While disparities resulting from local control remain largely immune to federal court challenge under\textit{ Rodriguez}, other disparities are not and may still be challenged.
3. Is the school finance system generally progressive or regressive with respect to assumed need factors?
That is, all else equal, do school districts and schools with greater shares of children in poverty, limited English proficient children, and children with disabilities have more resources per pupil than districts with fewer of these children? More broadly, are resource levels higher in those schools and districts that by virtue of their student populations, labor market conditions, and other factors outside control of school officials have higher costs of achieving educational outcomes. Do those who need more receive more?14

4. Does the school finance system provide enough resources such that children in those schools where costs of educational outcomes are higher have equal and/or sufficient opportunity to achieve desired outcome levels?

Table 14.1 summarizes these questions in relation to conventional and modified school finance terminology. The first question above and part of the second are addressed by Standard One, drawing primarily on the work of early legal theorists including Wise (1968) and Coons, Clune, and Sugarman (1969). This baseline standard asks first whether there is variation in resources that might be considered an equal protection violation and second whether that variation is specifically related to issues such as state granted local authority to raise supplemental per-pupil revenues.15

The second standard in Table 14.1 is the pure vertical equity or equal outcomes standard. Under the second standard, outcome equality is purely relative. For example, if the "average" student in a state receives a quality of education sufficient to provide a 25 percent chance of achieving an adequate educational outcome, so too should the child from economically disadvantaged background, the child with limited English proficiency, and the child with disabilities.16 Whether a 25 percent chance is constitutionally adequate is addressed under the third and fourth standards.

Pure vertical equity can be accomplished in a system that is adequate for none (according to judicial or legislative interpretation of state constitutional mandate). That is, pure vertical equity can be achieved when all children are deprived entirely of schooling. However, pure vertical equity can also be violated in a system that is adequate for all. As such, it makes practical sense to couple vertical equity and adequacy conceptions.17

Standard three provides an integrated conception of vertical equity and adequacy, which might also be framed as an adequacy compromise view of vertical equity. The vertical equity as adequacy standard is perhaps the most common approach in modern school finance "adequacy" litigation under state education clauses. This standard provides the basis for arguments that a school finance system is both generally inadequate and that the inadequacies are distributed inequitably. The primary emphasis under the vertical equity as adequacy standard is that children who would otherwise have less likelihood of achieving adequate educational outcomes require sufficient additional resources to increase their opportunities. Deprivation, which is most often unequally distributed with respect to student needs, must be eliminated. From a practical perspective, opportunity surpluses are viewed as nonoffensive under this standard.

Table 14.1 Finance Equity Synthesis Framework

<table>
<thead>
<tr>
<th>Level/Standard</th>
<th>School finance conception</th>
<th>Central question</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>Horizontal Equity &amp; Fiscal Neutrality</td>
<td>Are there differences in resources unrelated to educational need?</td>
</tr>
<tr>
<td>Two</td>
<td>Pure Vertical Equity (equal outcomes)</td>
<td>Are there sufficient differences in resources to accommodate educational need, measured against equity of outcomes standard?</td>
</tr>
<tr>
<td>Three</td>
<td>Vertical Equity as Adequacy (equitable and/or sufficient opportunity to achieve minimum outcomes)</td>
<td>Do all groups of children have sufficient resources to support equal opportunity to achieve minimum outcome standards?</td>
</tr>
<tr>
<td>Four</td>
<td>Adequacy</td>
<td>Is aggregate funding sufficient for children to achieve minimum outcome standards?</td>
</tr>
</tbody>
</table>

This third standard is likely the most immediately practical standard in modern school finance litigation for making marginal improvements to pure vertical equity. However, this standard can only legitimately improve
vertical equity where the outcome standard is sufficiently high. Where the outcome standard is very low, the state’s most vulnerable children will be guaranteed only sufficient opportunity to achieve a meaningless quality of education, while others obtain far more and are better positioned for access to higher education and the work force. While opportunity deficits are erased by lowering the outcome standard, the magnitude of opportunity surpluses increases.

Under the third standard, vertical equity is achieved only where all children have access to adequate opportunities whereas pure vertical equity can be maintained with this standard but only if no children have surpluses beyond this level. Typically, however, pure vertical equity is compromised under the third standard by the existence of opportunity surpluses.

The fourth and final standard is the pure adequacy standard, decoupled from vertical and horizontal equity. This standard asks broadly whether a state legislature has allocated sufficient funding statewide for a constitutionally sufficient share of the state’s children to achieve a constitutionally adequate educational outcome. Under a pure adequacy standard, the distribution—across districts, schools, and children—of either educational inputs or educational outcomes is largely irrelevant; such a standard may violate either or both principles of horizontal and vertical equity so long as a sufficient amount of funding is allocated for a sufficient share of the state’s children to succeed.

Directions for Future Research and Development

We have presented in this chapter the perspective that the shift in interest from educational inputs toward educational outcomes adds clarity to conceptions and measurement of vertical equity. We have suggested that the Who? and How much? questions of vertical equity analysis may be addressed by measuring, to the extent currently available data and methodologies permit, the marginal costs of achieving specific educational outcomes across students and settings. Further, we have suggested that such measures might provide the basis for future legal challenges centered primarily on vertical equity conceptions. Yet the shift in emphasis from financial inputs toward educational outcomes requires that we now pay much greater attention to the desired outcomes of schooling in America and how we measure those outcomes. (See Jacobsen and Rothstein, this volume, for a discussion of educational goals in the United States.) Alternatively, the current policy focus on educational outcomes necessitates a rethinking of school finance.

Numerous researchers since Garms and Smith (1970) have attempted to estimate statistically how the cost of achieving specific educational outcomes varies across individual students, groups of students, and educational settings. Invariably, in the current context of state standards and accountability systems, those studies have based their models on state tests of academic achievement, typically including only reading and mathematics achievement. The literature on the extent to which state reading and math tests have predictive validity toward labor market outcomes or even postsecondary outcomes remains sparse. Further, while courts have repeatedly emphasized civic participation as a critical goal of public schooling, few states include assessments of civic knowledge in their testing and little is known about the relationship between current testing and civic outcomes. The void between broad judicial interpretations of state constitutional mandates with respect to educational outcomes and how we measure those outcomes in practical context and empirical models is vast.

The accelerated movement toward state adoption of Common Core Standards and aligned curriculum and assessments may provide future opportunities to both evaluate costs of achieving broader outcome standards specified in the Common Core, and to more accurately compare costs of achieving common standards across states. Such cost comparisons across states have thus far been infeasible due to vast variations in the rigor of state standards and assessments, even across the same content areas.

It stands to reason that the Who? and How much? questions of vertical equity analysis might be answered differently in any given state under different outcome objectives. Where college matriculation and completion are the outcome of interest, marginal costs of achieving those outcomes across children and settings may differ from marginal costs of closing test score achievement gaps. The new emphasis on vertical equity, either in pure form or with respect to minimally adequate outcomes, requires far greater attention to the desired outcomes of schooling and how we measure them.

Others have presented a compelling argument that the overemphasis on measurable student achievement
outcomes in math and reading has led to an excessive narrowing of educational programming for children requiring the greatest effort to improve those specific outcomes (Rothstein, 2003). Where minimally adequate educational outcomes are defined only in terms of test scores in reading and math, broader, richer educational opportunities are reserved for a privileged few for whom the minimally adequate reading and math test scores are but an incidental rite of passage. As discussed above, we should seek to broaden outcome measurement to reduce the extent to which such problems occur. However, inequities in the breadth or depth of curriculum are unlikely to be remedied by simply adding outcome assessments for each and every possible course of study from Advanced Calculus to Advanced Jazz Improvisation. More than likely, effectively evaluating inequities in the breadth of curricular opportunities will require continued focus on educational inputs, as well as outcomes.

Notes

1. Hawaii’s school system is often described as a single school district. Baker and Thomas (2006) describe the system as direct state control over schools, where only the state legislature has taxing and revenue-raising authority for those schools.

2. Often only that level of authority is granted directly by state legislative action. Increasingly, school districts are being encouraged to decentralize control over the expenditure side of financial management, allocating lump sums to individual schools through weighted student formulas and granting increased budget authority to those schools. Even in the most aggressively decentralized forms of this model, as implemented in Seattle, personnel contracts remain managed at the district level and the total budget constraint is dictated by district level available revenue. For more extensive discussion and analysis see Baker and Thomas (2006).

3. See also Yinger (2004).

4. In New York State, for example, a cross-district distributional analysis would fail to capture disparities that may exist across more than one-third of the state’s children because they all attend schools within a single district. Hawaii has generally been exempted from distributional analysis despite significant inequities in financial resources and teacher quality across Hawaii schools (Baker & Thomas, 2006).

5. Exceptions include Stiefel, Rubenstein, and Berne (1998).


7. For example, measures that evaluate spending at the top and bottom of the distribution only (ranges and range ratios), the pattern of distribution across all districts (coefficients of variation, Gini coefficients), or distribution just among those districts in the lower half of the distribution (McLoone indices).

8. The two are not entirely irreconcilable. Adam Smith argued that individuals “ought to contribute to the support of the government, as nearly as possible, according to their respective abilities [ability-to-pay taxation]; that is, in proportion to the revenue which they respectively enjoy under the protection of the state [benefit taxation]” (Smith, 1776, p. 310, in Steuerle, 1999).

9. While on the one hand, equal protection challenges address in a purely relative sense the different treatment of otherwise similarly situated individuals, courts commonly evaluate differential treatment in terms of deprivation with respect to some minimum standard, as in deprivation of the right to vote. Unlike the right to vote, it is much more difficult to discern at what point along the education quality continuum deprivation has occurred.


12. District court judge Terry Bullock in Montoy v. Kansas provided one particularly bold proclamation to this effect. "In defense, Defendants simply argue 'money doesn't matter.' Without regard to the constitutional mandate that there be adequate funds for a suitable education and that those funds be equitably divided, the defense seems to say: there is no correlation between spending and student learning, so what’s all the fuss. 'Money doesn’t matter?' That dog won’t hunt in Dodge City!" Montoy II, 2003 WL 22902963.

13. Except in the special case where that outcome is “nothing” or “0.” The unique feature of this case is that the theoretical cost of achieving nothing is nothing. And, that cost does not vary by students or context.

14. For a comprehensive view on how one might measure with relative precision whether resources are sufficient to achieve desired outcomes, see the chapter in this volume by Duncombe, Nguyen-Hoang, and Yinger.
Notably, most concerns addressed under this standard cannot be addressed in federal court as a violation of equal protection, but may be addressed under state equal protection clauses.

This assumes the student’s IEP states that the student is capable of achieving a state standard.

Conceptually, the Equal Educational Opportunities Act (EEOA) and the Individuals with Disabilities Education Act (IDEA) provide statutory frameworks for advocating pure vertical equity. That is, one should be able to apply IDEA to advocate for educational opportunities for children with disabilities to be similar to educational opportunities for other children in the state, no more, no less. However, practical application of IDEA tends to be linked to external standards of “adequate” services for children with disabilities, but only for children with disabilities. This creates a legal protection imbalance across student groups that potentially compromises pure vertical equity. Pure vertical equity standards may also be useful for state level judicial review under equal protection clauses in states where state constitutional adequacy concerns are nonjusticiable. These claims still face the uphill battle of convincing courts that an equal outcomes standard is practicable.

In the immediate aftermath of the Coleman Report (1966), Garms and Smith (1970), in pioneering work, used school-level data from New York State to estimate a crude, regression-based education cost function model with the goal of estimating the costs of achieving specific student outcome levels on standardized reading and math assessments, given a rich set of student background characteristics. The authors posit that “equality of educational opportunity exists when the average achievement of groups is roughly equal. This definition recognizes a duty of the public schools, as servants of society, to attempt to overcome environmental deficiencies that are not the fault of the individual students” (p. 305). That is, in 1970, Garms and Smith further proposed an outcome-based definition as the value system for measuring vertical equity, and proposed and applied an empirical model for identifying “who” requires differential treatment and “how much.”

References


Measuring Equity and Adequacy in School Finance

THOMAS A. DOWNES AND LEANNA STIEFEL

Introduction

Over the past 45 years, researchers have devoted significant effort to developing ways to measure two important goals of state school finance systems: the promotion of equity and, more recently, the provision of adequacy. Equity, as the term is traditionally used in the school finance literature, is a relative concept that is based on comparisons of inputs (often aggregated into a per-pupil spending measure across school districts). Thus, an equitable finance system is one that reduces to a “reasonable level” the disparity in per-pupil spending across a state’s districts. Adequacy of funding, in contrast, is an absolute concept that requires that spending reach a minimum threshold that provides sufficient spending to give students in each district an opportunity to meet state standards of performance. Thus, adequacy focuses only on the bottom part of the distribution of spending, with no attention to variations above the threshold needed for adequacy. Although general definitions are straightforward, quantifying measures of either the equity or the adequacy of a school finance system is a challenge.

Other chapters in this section present several considerations that arise in the design of school finance systems to address equity or adequacy concerns. For example, Baker and Green (this volume) develop the legal, economic, and school finance concepts of equity and adequacy, while we focus on how researchers and policy analysts have translated those concepts into quantitative measures. In addition, we do not discuss in any detail the specific ways to adjust the quantitative measures for differences in prices of inputs across school districts, the differential costs of educating students who have educational disadvantages, or the higher costs experienced by districts with exceptionally small enrollments. The various ways of measuring these cost differentials are developed by Duncombe, Nguyen-Hoang, and Yinger (this volume). All of the measures of equity and adequacy that we describe can be, and usually are, adjusted for such cost differentials. Nor do we discuss the specifics of how school finance formulas have evolved to address issues of equity and adequacy; instead we refer the reader to Picus, Goertz, and Odden (this volume). Our emphasis is on how conceptions of equity and adequacy are translated into specific numerical measures.

The chapter is organized as follows. In the first section, we discuss statistical measures of school finance equity. In the second section, we identify issues that cut across the four main methods for measuring adequacy, and we then present and analyze those four methods. The third section discusses issues of incentives that arise when adequacy adjustments are included in state intergovernmental aid formulas. In the final section, we highlight particular areas in need of further research.

Measuring School Finance Equity

As described by Baker and Green (this volume), equity concepts are often separated into horizontal and vertical equity. Horizontal equity refers to how well students who are similar in their characteristics are treated relative to one another. Vertical equity refers to the degree to which students who differ from others, due to educational disadvantages such as poverty or due to differing fiscal capacities of their school districts, are treated appropriately differently.

While the equity concepts are defined in terms of the treatment of individuals, school finance systems are
designed for districts not individuals. Thus the concepts are translated from the individual to the district level by focusing on averages across groups of individuals. The concept of fiscal neutrality offers one example of how the individual-based concept of vertical equity is translated into a district-based equity concept. In particular, an aid system is said to be fiscally neutral if differences in the fiscal capacity of districts do not result in systematic differences in average per-pupil spending (Louis, Jabine, & Gerstein, 2003).

**Horizontal Equity Measures**

In their early work, Berne and Stiefel (1984) identified 11 possible statistical measures to quantify the degree of horizontal equity (or variation) in spending per pupil across school districts within a state.² Different measures reflect different values (Berne & Stiefel, 1984, p. 23). One value reflects whether all districts, as opposed to only a subset of districts, are included in the measure. Another value is whether the measure places a heavy emphasis on the lowest-spending districts. Yet a third value gauges whether equal proportional changes in spending (for example, a 5 percent increase in every district’s per-pupil spending) leaves the equity of the system unchanged. Although such a change in spending would leave all districts (sorted from low to high) in the same relative position, spending in the higher-spending districts would rise by a larger absolute amount. Some equity measures remain unchanged after such proportional changes while others indicate that equity has declined.³ Measures that remain unchanged after proportional changes are appropriate for making comparisons across states whose levels of spending differ approximately by proportional amounts.

In particular, the range measures the difference in spending per pupil between the highest- and lowest-spending districts. The range is easy to understand but it ignores all districts in the middle of the distribution, is sensitive to extreme outliers, and indicates greater inequity with equal percentage increases in spending. A second related statistic, the federal range ratio, corrects some of the problems of the range. It measures the difference between per-pupil spending at 95th and the 5th percentile divided by the 5th percentile. In contrast to the range, the federal range ratio is not sensitive to extreme outliers and invariant to equal proportional changes in spending in all districts. If the goal of state policymakers is to constrain districts to small differences in spending per pupil, the federal range ratio serves as a good indicator of how well that goal is being achieved.

Another measure, the McLoone index, focuses attention on the bottom of the distribution. It compares the sum of actual spending in all districts that spend less per pupil than the median district to what total spending would be in those districts if their spending were brought up to the median. The higher is the resulting ratio, the less is the inequity. Thus, the McLoone index focuses on the bottom half of the distribution and ignores any districts whose spending is above the median. Implicit in its construction is the idea that spending by the upper half of the districts is irrelevant to achieving school finance equity.

Other measures, such as the coefficient of variation, the Gini coefficient, the Theil measure, and the standard deviation of the logarithm of spending per pupil are based on per-pupil spending in all districts.³ These measures do not change when there are equal proportional increases in all districts’ spending per pupil. Moreover, all of these measures can be decomposed into within (state) and across (state) contributions to inequity when districts from several states are compared.

Berne and Stiefel (1984) tracked the behavior of the eleven equity measures over eight years in Michigan and 11 years in New York State and found that four groups of measures could be formed based on similarity in the patterns among them. The range, restricted range and variance were in one group; the coefficient of variation, Gini coefficient, Theil measure, standard deviation of logarithms,³ and relative mean deviation were in a second; the federal range ratio and the McLoone index each formed their own third and fourth groups.⁵ Thus analysts who wish to provide a comprehensive picture of the equity of a school finance system need choose only one measure from each group.

**Vertical Equity Measures**

To quantify vertical equity and fiscal neutrality, Berne and Stiefel developed two approaches. One approach uses regression analysis to relate spending per pupil at the district level to factors that relate to cost
differentials (such as the percent of students in poverty) or, in the case of fiscal neutrality, to property wealth per pupil. Higher coefficients on the cost factors imply that spending is higher in districts with above-average proportions of costly-to-educate students and, hence, that the interdistrict pattern is more vertically equitable than if the estimated coefficients were lower. In the equation to test for fiscal neutrality, small coefficients on the wealth variable imply that per-pupil spending in wealthy districts is not much higher than in poor districts and, hence, that the finance system is relatively equitable along that dimension.

The second approach converts per-pupil spending into spending per weighted student, where the weights are based on the differential costs associated with each type of student compared to students with average costs. The horizontal equity measures based on weighted students can then be used to assess the degree of vertical equity.

The importance of the choices that researchers make when choosing a particular statistic and the measures of a particular equity concept are highlighted by some of the differences in equity measurement used in two much-publicized rankings of state school finance systems: one produced by Education Week as part of its Quality Counts report, and the other generated by the Education Law Center in New Jersey. The 2013 Education Week rankings use per-pupil state and local revenues or expenditures (for 2009 or 2010), adjusted for regional cost variations, with students weighted extra for poverty (1.2 weight) and disabilities (1.9 weight). Four different equity measures are calculated (wealth neutrality, McLoone index, coefficient of variation, and restricted range). Florida is most equitable according to the McLoone index, West Virginia according to the coefficient of variation, and Utah based on the restricted range.

The Education Law Center (Baker, Sciarra, & Farrie, 2012) uses only state plus local revenues (from 2009), adjusts for regional costs, economies of scale, and population density, and calculates only one equity measure from simulations of whether states provide more or less funding to districts as poverty concentration increases. Utah, New Jersey, and Ohio rank first as most equitable.

Both Education Week and the Education Law Center are measuring versions of vertical equity (one with weights for poverty and disabilities and the other by comparing spending in high- and low-poverty districts) but they reach different conclusions in part because of the differences in weights and in part because of differences in the equity statistics they use. For example, Ohio provides relatively more funds for higher compared to lower-poverty districts and thus ranks high on the Education Law Center equity measure, but the higher-poverty districts often spend above the median for the state and thus Ohio does not have a particularly high McLoone index and does not rank at the top on the Education Week rankings.

Measuring School Finance Adequacy

The spending that is deemed adequate in a particular school district in a particular state depends upon the context in that state. The state’s constitution and its interpretation by the state’s courts and elected leaders ultimately determine what is adequate (Briffault, 2007). In some states, such as New Jersey and Kentucky, the courts have been relatively prescriptive in translating the language of the state constitution into specific adequacy standards and in specifying the steps policymakers need to take to meet those standards. In other states, including Alabama and Ohio, the courts have given policymakers more discretion both to establish the standards and to choose the policies designed to ensure that all students have access to an adequate education (Briffault, 2007). In yet other states, the courts have not been involved at all; instead legislators and governors have designed school finance systems that aim to achieve adequacy.

The goal of any method of calculating adequate spending is to determine the "cost" of achieving a specified level of student performance—that is, the least amount of per-pupil spending needed to achieve that level of student performance. The main approaches for "costing-out" a specified level of student performance are the professional judgment or resource cost model approach, the successful-districts approach, the evidence-based approach, and the district cost-function approach. Given the common goal of these approaches, all explicitly or implicitly build on the idea of a “production function for education” or an efficient relationship between student outcomes and spending. When the goal is calculation of adequate spending, researchers reverse this production function relationship; that is, they relate spending to outcomes instead of relating outcomes to

221
spending. A multiplicity of approaches have been developed to calculate adequate spending because the technology that links school inputs to student achievement is not well understood.

Most state policymakers set standards establishing levels of academic performance (or outcomes) that the students in each district are expected to attain. Differences across districts in input prices, student needs, and district characteristics (i.e., cost differentials) mean some districts will need to spend more than others to achieve the standards (Duncombe & Lukemeyer, 2002). Thus, the process of calculating adequate levels of spending usually is undertaken in two steps regardless of the methodology that underlies the actual calculations (Duncombe, 2002). In the first step, researchers determine the spending level needed to attain the state-established student performance requirements in at least one district, often labeled a benchmark district. In the second step, researchers adjust the spending in the benchmark district to reflect the cost differentials present in other districts.

Professional Judgment/Resource Cost Model Approach

In the standard application of the professional judgment approach, researchers consult with professional educators ("experts") to decide the level of spending per pupil that is required to achieve an adequacy standard in a prototypical school with predefined characteristics. Examples of characteristics generally include the total enrollment in the school, the percentage of students who are poor, the percentage who are English language learners, and the percentage of students with individual education programs. The experts provide detailed information on the inputs needed to achieve the adequacy standard, based on their best judgments from their own experiences, and researchers combine estimates from panels of such experts to arrive at "best practice" or benchmark schools and districts. Spending for the benchmark district (or districts) is then adjusted for cost differentials, typically though the use of input price and cost indexes.

The quality of the estimates generated by this method is based heavily on the process involved. Panels of experts generally meet for several days allowing time for detailed discussion and debate about what is needed to achieve predefined standards. Their calculations are often passed to other panels that look at several estimates all together and choose which ones to recommend. Sometimes the original panels are asked to reconvene after all have done their work to reconcile their estimates with those of others.

A major problem with the professional judgment approach is that participants in the panels are not asked to contemplate trade-offs, which means that they tend to adopt a pie-in-the-sky view of the world. Sonstelie (2001), Rose, Sonstelie, Reinhard, and Heng (2003), and Rose, Sonstelie, and Richardson (2004) suggest that this problem can be avoided if a two-step approach is used, with professionals first being asked to make resource allocation choices given a series of different budgets and then being asked to indicate which of the budgets would enable them to satisfy certain standards. To support this approach, Rose, Sonstelie, and Richardson (2004) document cases in which, given explicit outcome standards, the same set of professionals made different resource recommendations when they did and did not face explicit budget constraints.

Finally, the professional judgment approach suffers from two forms of human bias: self-serving behavior and habit. Some of the experts may have their own reasons for wanting funding to be high or low and thus may make recommendations that serve their own purposes but not necessarily the purpose of the project. While the remaining experts on the panel could counter this bias, the absence of a budget constraint substantially reduces the incentive for any participants to argue strenuously for compromises. In addition, the experts are using their experience of current classrooms, schools, and districts, which may or may not be a good base for understanding what is required to meet state standards. If the professionals are given a standard that few, if any, districts are attaining, then their estimates of the resources needed are not likely to be accurate.

Several of the existing professional judgment studies highlight the extent to which the results are dependent on personal experience. For example, in a report that uses the professional judgment methodology to cost out an adequate education in Nebraska, Augenblick and Myers (2003) include a table that summarizes, for elementary schools in a relatively large school district, the personnel recommendations generated by professional judgment panels in Nebraska, Kansas, Maryland, and Montana. The extent of variation in the recommendations is striking, even though contexts were intended to be very similar. Because the professional judgment method typically imposes constraints on class size, the recommended number of teachers did not
vary dramatically across the states. But the number of teacher aides per 1,000 students ranged from zero in Maryland to 25.7 in Nebraska. The number of librarians and media specialists ranged from two in Maryland to five in Kansas, and the guidance counselors ranged from 1.4 in Nebraska to five in Kansas. Total personnel recommended ranged from 80 in Maryland to 102.9 in Nebraska.

While in theory these differences could reflect differences in state standards, in this case, the standards used by Augenblick and Myers (2002, 2003) for Montana and Nebraska were very similar, as were the characteristics of the prototype schools and the students they served. Further, in Montana, the output standards presented to the professional judgment panels were based on statewide performance results, not on district-specific performance information. As a result, professionals with experience in high-performing or low-performing schools or districts would have little, if any, experience on which to draw to determine the resources needed for a prototypical school to achieve the performance gains specified in the standard.

**Successful-Districts Approach**

The successful-districts approach builds on the plausible idea that districts already meeting a state’s performance standard will be spending an amount that is at least sufficient to provide an adequate education. To implement this method, researchers identify districts (not schools) presently meeting the standard and then measure how much these districts are spending. To mitigate the influence of extraordinary cases and to reduce the possibility that districts with inefficiently high spending affect adequacy estimates, most practitioners eliminate an arbitrary percentage of outlier districts (those that spend the least and the most). For the remaining benchmark districts, average (either mean or median) per-pupil spending is calculated. Finally, to determine what spending is needed to meet the standard in districts outside the benchmark group, per-pupil spending is adjusted for differences in such determinants of costs as the size of the districts, the types of students served, and the salaries of teachers.

Basing adequacy calculations on the spending of districts that have already attained the standard has considerable intuitive appeal. The successful-districts approach, however, may fail to produce an accurate estimate because districts that have attained the standard may not be representative of “typical” districts, particularly if the standard is high. In other words, the districts that are meeting standards are not a random sample of all districts. They may differ in terms of measurable characteristics, such as their wealth or the types of students they serve, and in unobservable ways that are related to the cost of achieving the standard. Duncombe and Lukemeyer (2002) document that in New York State the extent to which successful districts are atypical in measurable ways, both in terms of property wealth and the share of high-needs students, increases as the standard increases. While researchers have implemented various ad hoc approaches to account for the fact that successful districts represent a select group, they have not agreed on a standard method for addressing this limitation.

**Evidence-Based Approach**

One of the earliest methods to determine the cost of providing an adequate education was the whole-school design approach (Odden, 1997; Odden & Busch, 1998). Proponents of this method argued that experience from successful school reform efforts can be used to determine the expenditures needed to provide an adequate education. While the method was never widely applied, possibly due to the “black box” nature of whole-school reforms and the select nature of those settings in which whole-school reforms are successful, the logic of using research results to guide the costing-out process morphed into the evidence-based approach.

Applications of the evidence-based approach start by first convening a panel of researchers to produce a report that summarizes the relevant educational research. Where possible, experimental evidence is used to identify successful strategies, but the researchers also draw on evidence from more traditional statistical analyses of educational strategies and from evaluations of successful whole school reform efforts (see, for example, Picus et al., 2013). Once a menu of strategies is developed, the researchers determine the cost of these strategies and then calculate the cost of operating a school or a district that adopts a mix of strategies chosen to ensure adequacy. Advocates of the approach argue that it has the advantage of providing educators with
direction on what to do while continuing to provide considerable leeway to choose strategies appropriate to local contexts.

In addition to building on the whole school design approach, the evidence-based approach is a natural extension of variants of the professional judgment approach, advocated for by Guthrie and Rothstein (1999) and the American Institutes for Research and Management Analysis and Planning (AIR/MAP) team of Chambers, Guthrie, and Smith (American Institutes for Research and Management Analysis and Planning, Inc., 2004), in which research is used to guide the deliberations of the panels. In New York in 2004, for example, panels were provided with "an objective description of mainstream educational research as background for [their] deliberations." In principle, providing panel members with a summary of the consensus in the literature could constrain the panel deliberations in reasonable ways. Panel members might feel a need to justify any deviations between the resource mixes they propose and those that the educational research literature suggests are effective.

In reality, however, because of the lack of consensus in the literature on the strength of the links between school inputs, such as class size and teacher characteristics, and student performance, any "description of mainstream educational research" is likely to be either uninformative or colored by the biases of the researcher preparing that description. Moreover, experimental evidence or even rigorous statistical analysis of the role of many inputs or programmatic reforms in improving student performance is lacking or just being developed. In addition, and related in terms of validity, the estimate of adequacy is often based on evidence from a highly select sample.

**School District Cost-Function Approach**

Researchers who use the cost-function approach apply economic theory to understand the behavior of local governments and to explain variation in expenditures across school districts. Specifically, these researchers use as their starting point the concept of a production function for education, in which educational outcomes (Q) are a function of inputs (X). Implicit in this production relationship is a cost function:

\[ C = pX = c(Q, p, A) \]

where

- \( C \) is cost per pupil
- \( p \) consists of input prices
- \( X \) consists of inputs to the production process
- \( Q \) consists of outputs of the production process
- \( A \) consists of variables that measure attributes of the school district and its students that influence its costs
- \( c(.) \) is the functional form relating costs to its determinants. 12 The cost function gives the cost of obtaining each set of outputs, conditional on input prices and district and student characteristics.

When asked to calculate adequate spending, economists typically turn to this method (see, for example, Imazeki & Reschovsky, 2006). When advising policymakers, however, researchers have been reluctant to recommend the use of this method because it is both difficult to explain and not transparent.13

With this method, estimates of the extent to which certain factors contribute to cost variation is determined by relating variation in per-pupil expenditures to variation in output levels, input prices, student characteristics, and district characteristics. Information for step one (establishing a benchmark district) and step two (adjusting the benchmark to reflect cost differentials) is estimated together. As examples, outputs may include student performance standards such as test scores or graduation rates; input prices may include teacher and other personnel salaries; student characteristics may include percentages of students who are poor, English language learners, and disabled; and district characteristics may include size, density, and perhaps rural or urban location. In addition, in order for these models to represent cost functions, a measure of district efficiency should be included since costs represent minimum per-pupil spending for given output levels while per-pupil expenditure, which is the dependent variable used in the empirical analysis, represents actual per-
pupil spending for given output levels (see Duncombe, Nguyen-Hoang, & Yinger, this volume, for further explication). Researchers use econometric methods and historical data on school district expenditures, outcomes, and student and district characteristics to estimate a cost function. For a summary of this literature, see Golebiewski (2011).

Duncombe (2002), for example, uses this approach in a study of New York State. In estimating the cost function, he measures student performance in each district by scores on fourth- and eighth-grade math and English tests as well as state high school exams in math and English (Regents’ exams). After generating estimates of the cost function, he calculates how much a district with the mean value of each determinant of costs (a benchmark district) would have to spend to reach a particular performance standard. Adequate spending in other districts is calculated by multiplying spending in the benchmark district by a cost index, which has been calculated for each district using the cost-function estimates.

This approach has the virtue of estimating one equation that can incorporate different output standards and different student characteristics, should these factors change in the future. In addition, because the estimates are generated using statistical procedures, analysts can calculate confidence intervals for adequacy numbers. There are, however, at least four problems with this approach. First, the method requires high-quality data. If the data are imprecisely measured or incomplete, generating accurate results will be difficult. For example, in theory outputs should be measures of value-added or gains in performance from one year in a grade to the next year in the next grade. However, while some performance measures, such as performance on the ACT or the SAT, no previous score is available. As a result, performance on a nonaligned test must be substituted. As another example, poverty of students is generally represented by the percentage of students receiving free or reduced price lunch, but this may not be an accurate measure of student deprivation, especially at the high school level, where students are reluctant to sign up for the federal lunch program even when they are eligible.

Second, measuring the efficiency with which districts deliver education is not straightforward and while researchers have tried several methods, such as indexes based on data envelopment analysis or indexes representing the degree of competition facing each district, still no widely accepted method exists. For example, Duncombe (2002) includes in his cost function an explicit measure of inefficiency in resource use, based on a data envelopment analysis (DEA). Each district’s cost index is calculated with that district’s efficiency measure replaced with the average level of efficiency. Downes and Pogue (1994) also outline methods that could be used to generate cost indexes. They argue that the fixed-effects estimation methodology they use could, indirectly, account for some unobserved determinants of costs, unobserved outputs, and inefficiency. Duncombe and Yinger (2011) argue for including measures of community characteristics that could influence the extent of a district’s efficiency. (See Duncombe, Ruggiero, & Yinger, 1996, and Imazeki & Reschovsky, 2005, for more detail and other approaches.)

Third, theory does not indicate a specific “functional form” for the regression model, and researchers obtain different results depending on the particular functional form they use. For example, in Texas two groups of researchers obtained different results in part because they used different functional forms. One group used a modified Cobb-Douglas specification, which generated a spending equation that was linear in logarithms, while the other used a trans-log specification, resulting in an equation that involved many interactions of the logarithms of inputs and prices on the right-hand side. Although these differences are quite technical, each specification has its origins in economic theory about organizational behavior, and, theoretically, one is not more correct than the other (Imazeki & Reschovsky, 2005; Gronberg, Jansen, & Taylor, 2011).

Finally, the method has a “black box” quality to it. This means that although the cost-function methodology allows researchers to determine an overall level of spending needed for adequacy, the methodology does not specify how resources should be allocated to produce the standard desired.

Using Equity/Adequacy Measures to Design School Finance Systems

The amount each district would need to spend to provide an adequate education is a critical component of
most state or federal intergovernmental aid formulas (Sonstelie, 2004). As a result, methods used to determine
the adequate level of spending should explicitly address certain design features that have been discussed in the
intergovernmental aid literature. This section contains a brief discussion of these design considerations; more
expansive discussion of design considerations can be found in Louis, Jabine, and Gerstein (2003).

Before turning to these specific design features, we mention again one pervasive reality common to all
methods of estimating adequate spending amounts: because in most states numerous districts are far from
meeting existing standards, all costing-out methodologies must project “outside” the information currently
available for at least some districts. Since this problem is unavoidable, researchers would do well to make
explicit the inherent unreliability of the adequacy calculations for some districts. Checks on reliability that we
discuss below could help reduce, but not eliminate, some of the lack of reliability.

**The Incentive Problem**

In the intergovernmental aid literature, broad consensus exists that state aid should not compensate local
jurisdictions for factors that can be influenced by the behavior of the governments receiving aid (Louis, Jabine,
& Gerstein, 2003). For example, in the literature on aid to provide compensation for the costs associated with
special education, much attention has been paid to the incentive that drives school districts to identify too
many students as requiring services. This incentive occurs when an aid formula compensates districts for the
number of students that the district itself identified as needing special education services (Cullen, 2003). Many
states have responded to the existence of these problematic incentives by modifying their aid formulas. Under
the modified formulas, the aid a district receives does not depend directly on the number of students that the
district has identified as needing special education services. As a result, some districts may not be compensated
fully for the cost of serving students needing special education services.

In theory, then, the calculation of the amount each district would need to spend in order to provide adequate
services should be independent of the choices that a district makes. In practice, breaking the link between aid
and district choices may be very difficult to do, since many of the factors that account for cross-district
variation in district costs are also affected by district choices. A commonly cited example is district size, which
depends on decisions to pursue consolidation. District size is not the only cost factor that could create
problems; some observers argue that even such cost factors as the fraction of students eligible for school lunch
subsidies can be influenced by how aggressively districts seek to identify eligible students.

Downes and Pogue (1994) and Duncombe (2002) argue that one strategy state policymakers can pursue in
order to avoid the effects of adverse incentives of this type when designing aid formulas is to make cost
adjustments dependent on regional or statewide averages rather than the actual measures of a district’s
discretionary factors. While this approach would mitigate the incentive problem, since district behavior can
have little or no influence on regional or statewide averages, the approach also leads to incomplete adjustment
for cost variation.

A second alternative is to turn to the research literature for evidence on the extent to which district behavior
is sensitive to these incentives. The research by Lankford and Wyckoff (1996) on the extent to which special
education assignments by districts in New York were sensitive to the incentives in the aid formula offers a nice
example of the kind of research that can guide policymakers. If research indicates that district behavior is not
particularly responsive to these incentives, then full compensation is warranted. Further research on
responsiveness of school districts to incentives will help improve the ability of policymakers to design aid
formulas that compensate for cost variation without providing perverse incentives.

**Updating the Adequacy Calculations**

Almost any method that could be used to determine adequate spending levels will be time-consuming and
potentially costly. The natural inclination will be to use the results of the initial study as the basis of aid
calculations for several years, with the only substantive year-to-year changes being inflation adjustments that
are uniform across the state. Following this natural inclination could, however, lead to lack of adequate
funding for many districts.
Over time, new data will become available, characteristics of districts will take on different values, and simple inflation adjustments will not account for these changes. Since most methods use a process that can incorporate new data on characteristics, recalculating adequate spending amounts can be done easily no matter which method is used. For example, new counts of students in poverty can be inserted into indexes that adjust for student cost differences without reestimating the entire cost function. As a result, the initial adequacy study will not need to be fully replicated on a yearly basis, only updated. Recent work by Augenblick, Palaich, and Associates (2013) on Colorado offers an example of how new adequacy calculations can update, rather than redo, previous costing-out studies.

State standards are also likely to evolve. Depending on the method used to determine the spending levels needed to satisfy the standards, adjusting for changes in these standards may be difficult. If, as is the case with the cost-function methodology, the method incorporates information on student performance without being dependent on a specific performance standard, a change in standards can be handled as easily as can new data. If, as is the case with the professional judgment methodology, spending in the benchmark district is determined in reference to specific standards, then a change in the standards necessitates a new adequacy study.

## Needed Further Research

Eric Hanushek (2005, 2007) has questioned the scientific underpinnings of the methods used to calculate the cost of an adequate education. Most of the criticisms made by Hanushek echo concerns noted above. But Hanushek goes further, contending that these concerns are of sufficient magnitude to render meaningless any adequacy calculation based on existing methodologies. He concludes that:

> Decisions on how much to spend on education are not scientific questions, and they cannot be answered with methods that effectively rule out all discussions of reforms that might make the school system more efficient.

(Hanushek, 2005, p. 73)

Duncombe (2006) has responded to Hanushek’s criticism by arguing that, while the political process should determine the achievement and equity standards in each state, technical analysis using one of the methodologies described above should be the basis for determining the dollars that should flow to each school district. If the methodologies are flawed, researchers should work to eliminate those flaws. To do this, and to check the scientific merit of the methodologies, researchers utilizing any of the methodologies should evaluate the reliability and accuracy of their results, Duncombe (2006) argues. He suggests steps that could be used to evaluate the reliability, statistical conclusion validity, construct validity, predictive validity, and internal validity of the results. One critical task for researchers is to execute these steps and then to use the results of the reliability and validity checks to improve the methodologies.

Duncombe (2006) focuses on strategies that could be applied within the confines of a single methodology. Since in several states multiple methodologies have been applied, this situation provides researchers with an opportunity to determine the consistency of the results from different methodologies. Downes (2004) provides an example of relatively crude consistency checks; Baker (2006) offers a model for more sophisticated consistency checks. Such consistency checks may help to highlight correctable flaws in a methodology and also to illustrate the degree of imprecision involved in determining adequate funding.

Another fruitful avenue for research is to explore the benefit of combining one or more of the methodologies. For example, such a hybrid method could combine elements of the successful-schools, cost-function, and professional-judgment methodologies (Downes, 2004; Bradbury, 2004). Many recent studies implement multiple methodologies (see, for example, Augenblick, Palaich, & Associates, 2013), but few use the multiple methodologies in this hybrid manner. An exception is the Education Policy Improvement Center’s “Washington Adequacy Funding Study” (2007). In this analysis, the successful schools methodology was used to identify schools that appeared to be meeting the standard with little waste (inefficiency). Surveys of administrators in these schools were used to develop estimates of the amount of spending each school would need in order to provide an adequate education. Then, following the logic of Rose, Sonstelie, and Richardson (2004), these estimates were used to generate baseline budgets that were a critical input in a professional judgment study. The professionals were given these baseline budgets and a set of strategies chosen using an
evidence-based approach and were asked to choose the strategies needed to achieve the state standards and to make and justify any expenditures beyond baseline.

Combining the methodologies as was done in the Washington case mitigates the effect of the inherent flaw in the professional judgment approach, which is that it relies exclusively on people’s experience. Using the successful-districts methodology to select the sample provides a mechanism for avoiding the possibility that baseline budget estimates are flawed as a result of the inclusion of data from school districts that are operating inefficiently. Further experimentation with such hybrid methods will make it possible to determine if these methods work and if they provide us with any useful information about the validity of the individual methodologies.

Yet a further area for future research is the application of adequacy models to intra-district (cross-school) distributions of resources. School finance aid flows from states to districts and districts then make decisions about resources that students in each school receive. If districts aim to promote the achievement of state standards, logically the distribution of resources to schools would reflect the resources needed to allow students to reach adequacy. Could methods for determining adequate resource levels for districts be applied to schools as well? Such applications are not obviously appropriate. While it may be possible to estimate school-level equations that relate expenditures per pupil to student and school characteristics (such as the percent of pupils in poverty or school enrollment), one would need to make an argument for interpreting the estimated results as a cost function. Schools, like districts, are not necessarily cost-minimizing organizations. As a result, they may not operate efficiently. Thus, as with the district cost functions, problems arise because of the absence of agreed-upon ways of correcting for inefficiency. (See Schwartz, Stiefel, & Bel Hadj Amor, 2005, for some seminal work in this area.) In addition, while price differentials may exist across schools, it is not clear how to interpret these differences in those districts where teachers are all paid on one scale and the central district purchases most inputs for all schools.

Finally, although adequacy is currently a more prominent public goal than equity, many policymakers continue to believe that equity in school finance is important as well. (See Koski & Reich, 2006 for a philosophical/legal argument in favor of an equity emphasis.) Under certain circumstances, however, the two goals for school finance systems may be in conflict. For example, a low state standard for performance could lead a state to provide low levels of funding to meet an adequacy goal. At the same time, if the state does not restrict spending by wealthy districts from their own funds, then some standards of equity would be violated.

School finance analysts have been engaged in measuring equity of systems for well over 40 years, while the measurement of adequacy is more recent. There is much room for continued research on how to measure adequacy, especially on how to combine existing measures to capture the best features of each.

Notes

1. Recently, researchers have proposed that equity analyses use inputs known to be effective in improving student performance, specifically teacher value-added (Bastian, Henry, & Thompson, 2013).

2. The 11 measures are the range, restricted range, federal range ratio, relative mean deviation, McLoone index, variance, coefficient of variation, standard deviation of logarithms, Gini coefficient, Theil’s measure, and Atkinson’s index (p. 19).

3. This value is often labeled relative inequality aversion in the broader literature on measures of inequality.

4. The Gini coefficient is a measure based on the Lorenz curve, and the Theil coefficient is a measure based on information theory. The formulas for each are as follows:

   \[
   \text{The Gini coefficient: } \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} P_i |X_i - X_j|}{2(\sum P_i)^2 \overline{X}}
   \]

   \[
   \text{Theil’s measure: } \left( \frac{\sum_{i=1}^{N} P_i (X_i \log X_i - \overline{X} \log \overline{X})}{2 \overline{X} \sum P_i} \right) + \frac{1}{N} \sum_{i=1}^{N} P_i \left( \frac{X_i}{\overline{X}} \right) \left( \log \frac{X_i}{\overline{X}} \right)
   \]

5. More fully stated, the standard deviation of the logarithm of per-pupil spending and the relative mean deviation of per-pupil spending.

6. The Atkinson index changes groups depending on the value of one of its parameters, which can be set by the analyst in order to focus on the lower, the upper, or the middle of the distribution of spending per pupil.
7. Student counts are higher in districts with many weighted students, making their spending per (weighted) student lower.

8. *Education Week* combines the four equity measures with four spending measures and then gives each state a single school finance grade and rank. Here we focus on the equity part of these grades only.

9. The Education Law Center uses three other measures in its “fairness” rankings: funding level, funding effort, and coverage of public education.

10. Grubb (2006) argues that what is adequate depends on whether the courts and policymakers determine if the concept of adequacy is applied to access, to funding, to resources, or to outcomes. For example, if the concept is applied to access, then the standards that are relevant are minimum school standards, such as accreditation standards. If the adequacy concept is applied to outcomes, then the standards that are relevant are student performance standards established by state law. Many courts, such as the courts in Nevada and Montana, have applied the concept of adequacy to access and to outcomes. In such contexts, both minimum school standards and performance standards are relevant.

11. Often spending is limited to that devoted to “regular” education programs, with spending for special education or English language learners or compensatory help removed.

12. The functional form of a regression equation is the precise mathematical equation used by the researcher when that researcher estimates the regression equation (Schmidt, 2005).

13. Bradbury et al. (1984) represents an exception to this practice. The estimated cost adjustments provided in that paper were used in the construction of a formula to allocate a portion of the aid to cities and towns in Massachusetts.

14. Duncombe and Yinger (2011) discuss potential solutions to the first three of these problems in some detail.

15. Data imprecision is not a problem that is unique to the cost-function methodology; even the data that is generated by experts participating in professional judgment panels is likely to be imprecise. So any estimate of adequate spending is likely to be subject to error. In fact, as already pointed out, one of the virtues of the cost-function approach is that the methodology can provide a statistical estimate of the magnitude of the error.

16. DEA is an operations research technique used to determine whether an organization is producing the most of at least one of its outputs from its inputs. It is similar to input/output analysis and is explained more fully in Duncombe, Nguyen-Hoang, and Yinger (this volume).

17. Lankford and Wyckoff (1996) find what they describe as “at best weak support for the notion that New York school districts alter special education classifications in response to the financial incentives of the state aid formula” (pp. 238–239). In particular, for two of the three special education placement settings they consider, they find a no link between the dollar increase in excess-cost aid that results from an additional weighted special education student (holding total enrollment constant) and total special education enrollment. For the third placement setting, the elasticity of special education enrollment with respect to the dollar increase in special education aid is only 0.02.

References


finance and organizational structure in New York State schools (pp. 89–97). Albany, NY: Rockefeller College of Public Affairs and Policy, University at Albany, SUNY.


Measurement of Cost Differentials

WILLIAM D. DUNCOMBE, PHuong NGUYEN-HOANG, AND JOHN YINGER

Introduction

The evaluation of education cost differentials across school districts has been an important topic in education finance research for decades (Fowler & Monk, 2001). Interest in this topic has grown in recent years with the emergence of adequacy as the primary standard in school finance litigation and the growth of state accountability systems that focus on student performance. These developments call attention to education cost differentials, which arise when some districts must spend more than others to obtain the same performance. Despite a growing literature on the link between research and policy on this topic, however, existing state aid formulas usually contain ad hoc cost adjustments that fall far short of the across-district cost differences estimated by scholars. The objective of this chapter is to synthesize the research literature on education cost differences across school districts and to discuss the implications of this literature for state education aid formulas. The material in this chapter complements the discussions of equity and adequacy in the Baker and Green, and Downes and Stiefel, chapters in this volume.

The term “cost” in economics refers to the minimum spending required to produce a given level of output. Applied to education, cost represents the minimum spending required to bring students in a district up to a given performance level. Education costs can be affected by three categories of factors, each of which is outside of school district control: (1) geographic differences in resource prices; (2) district size; and (3) the special needs of some students. In this chapter, we address the principal methods for estimating the cost impacts of each of these factors. As discussed below, these impacts need not be the same, of course, for every measure of student performance or in every state.

While states commonly adjust their basic operating aid programs for differences in the capacity of school districts to raise revenue, typically measured by property wealth or income, few states systematically adjust these programs for cost differences across districts. Instead, cost adjustments tend to be confined to ancillary aid programs or to be nonexistent. In her recent survey, Verstegen (2011) finds that 16 states provide no extra funding for low-income students and 13 provide no extra funding for English language learners. This limitation is important because the success of a basic operating aid formula in providing the funds needed for an adequate education in each district, however defined, is linked to the accuracy of the cost adjustment. An aid formula cannot achieve an adequacy standard defined as a minimum level of student performance without adjustments for student needs, district size, and geographic variation in resource prices. Because cost adjustments may vary with the measure(s) of student performance or with conditions in a state, standard cost adjustments are not available. Instead, each state needs to estimate its own cost adjustments or else settle for approximate adjustments based on studies in similar states. Estimating cost adjustments is a challenging enterprise that requires clear judgments about state educational objectives, good data, and technical expertise. These challenges are the subject of this chapter.

The focus here is on cost differentials across school districts, not across individual schools. School districts are the primary budget decision making units; taxing power and budget authority also lie with district officials. Accordingly, state school finance systems are focused on distributing education aid to school districts, not schools, in most cases. As a result, this chapter does not cover recent research about inequity in the distribution of resources across schools within large urban school districts (Baker, 2012) or about weighted-student formulas designed to address this inequity (Baker, 2009; Ladd, 2008). Nevertheless, the measures of cost differentials due to student needs discussed in this chapter may also be appropriate for intradistrict funding.
The chapter is organized roughly in line with the major cost factors. We begin by discussing briefly the one method that can produce estimates for all three types of cost factors—education cost functions. We then turn to looking at other methods for estimating geographic resource price differences, the cost effects of enrollment size, and the cost impacts of various student characteristics. Each section describes the most frequently used methods, discusses their strengths and weaknesses, and provides key references for more detailed information.

**Education Cost Functions**

To estimate the relationship between spending, student performance, and other important characteristics of school districts, many education researchers employ one of the key tools of production theory in microeconomics, namely, a cost function. Cost is defined as the minimum spending required to reach a given level of student performance using current best practices. A district using best practices is said to be efficient. Cost cannot be directly observed, however, so cost functions are estimated using district spending (usually operating spending per pupil) as the dependent variable. Spending is higher than cost when school districts are inefficient—that is, when they deviate from current best practices. As a result, cost functions need to be estimated with controls for school district efficiency.

More formally, education costs, \( C \), depend on (1) student performance \( (S) \); (2) resource prices \( (W) \), such as teacher salaries; (3) enrollment size \( (N) \); and (4) student need measures \( (P) \), which are discussed in detail below; that is, \( C = f(S, W, N, P) \). Now let \( e \) stand for school district efficiency in delivering \( S \). Without loss of generality, we can set the value of \( e \) at 1.0 in an efficient district, so that it has a value between zero and one in a district that does not use current best practices. We then can write the cost/efficiency equation that scholars estimate as:

\[
E = \frac{C}{e} = \frac{f(S, W, N, P)}{e}
\]  
(1)

This formulation makes it clear that a district that does not use best practices \((e < 1)\) must spend more than an efficient district \((e = 1)\) to achieve the same level of performance \((S)\), all else equal.

Equation (1) has been widely used in various forms in empirical work because it addresses many fundamental questions of interest to scholars and policymakers. For example, a cost-function measures how much a given change in teacher salaries, district enrollment, or student needs affects the cost of achieving a particular level of student performance at a given level of efficiency. The cost-function methodology has been refined over the last few decades, and cost-function studies have been undertaken for several states.²

In order to estimate equation (1) using multiple regression analysis, researchers must address several methodological challenges. The first challenge is to identify a performance objective and find data to measure it \((S)\). One common approach, for example, is to select the performance measure or measures that are most central to a state’s school accountability system, which typically include student performance (measured by average scores or passing/proficiency rates) on state-administered tests in English, reading or mathematics, and perhaps graduation rates. In addition, these measures of student performance are determined simultaneously with district spending, so they need to be treated as endogenous when equation (1) is estimated.²

Together with a few state accountability systems, cost-function studies, for example Imazeki and Reschovsky (2006) and Gronberg et al. (2011), may focus not on levels of student performance but instead on the change in student performance over time, often referred to as a value-added measure. As shown in Duncombe and Yinger (2011b), this approach is difficult to implement in a cost study, however, because a value-added approach requires test score information on the same cohort in different grades—information that is not generally available. Moreover, value-added measures provide noisy signals about student performance, particularly in small school districts (Kane & Staiger, 2002).

A second methodological challenge is to control for school district efficiency \((e)\). The problem is that efficiency cannot be directly observed, so a researcher must select a method to control for efficiency indirectly.
Several approaches, each with limitations, have appeared in the literature. One approach is to estimate the cost function with district fixed effects, which control for all district characteristics, including efficiency, that do not vary over time (Downes & Pogue, 1994). The limitations of this approach are that it cannot control for district efficiency that varies over time and that, by removing all cross-section variation, it undermines a researcher’s ability to estimate the impact of \( S, W, N, \) and \( P \) on costs.

Another efficiency approach is to estimate a cost frontier based on the lowest observed district spending for obtaining any given student performance, to calculate each district’s deviation from this spending as an index of inefficiency, and then to control for this measure in an estimated cost function (Duncombe & Yinger, 2000; Reschovsky & Imazeki, 2003). A limitation of this approach is that this index of “inefficiency” reflects both cost and efficiency differences across districts. As a result, this approach may lead to underestimated coefficients of cost variables, such as student poverty, because a portion of the impact of these variables on costs may be captured by the estimated coefficient of the “inefficiency” index.

The final efficiency approach in the literature is to identify factors that have a conceptual link to efficiency and then to control for them in a cost-function regression. A limitation of this approach is that these conceptual links cannot be directly tested. Nevertheless, a strong case can be made for the inclusion of two types of efficiency controls. First, some district characteristics might influence the incentives for voters to monitor school officials or for school officials to adopt best practices. For example, Imazeki (2008) and Imazeki and Reschovsky (2004 and 2006) control for efficiency using a measure of competition from other public schools, which might influence the behavior of school officials. Second, some district characteristics, such as median household income or tax price, might influence voters’ demand for measures of school district performance other than \( S \). Because efficiency can be defined only relative to specific measures of \( S \), any spending to obtain other performance measures is, by definition, inefficient. Income and tax price are examples of variables that help control for this type of inefficiency (Baker, 2011; Duncombe & Yinger, 2011a and 2011b; Nguyen-Hoang & Yinger, 2014).

Because efficiency is defined by the selected performance measures, the division between costs and efficiency may be quite different in different states. States with different performance measures in their accountability systems—and hence in their cost analyses—should not expect to have the same variation in costs across districts. The factors influencing efficiency could also vary from state to state even if the performance measures are the same.

A third challenge is to select a functional form for the cost model. This form reflects underlying assumptions about the technology of production, such as the degree of substitution between inputs, economies of scale, and the interaction between school and nonschool factors. Most education cost studies have used a simple multiplicative cost function, which works well in practice but which imposes limits on both factor substitution and economies of scale. By contrast, Gronberg et al. (2011) use a flexible cost function that does not impose significant restrictions on production technology. This approach adds many variables to the cost model, however, which makes it more difficult to identify cost effects with precision.

The cost-function approach has been criticized by Costrell, Hanushek, and Loeb (2008) and Hanushek (2007). These scholars do not offer a better alternative to cost-function estimations, however, and their criticisms are addressed in Baker (2006), Duncombe (2006), and Duncombe and Yinger (2011b). Despite empirical challenges involved in estimating cost functions, therefore, they have some clear advantages over other methods of estimating cost differentials. First, they use actual historical data with appropriate statistical procedures to separate the impact of factors outside and within district control on the cost of reaching student performance levels. Second, they can provide measures of overall cost differentials across districts as well as measures of individual cost factors (resource prices, enrollment, and student needs) that can be used in state aid formulas. Some scholars have criticized cost functions on the grounds that their technical complexity makes them difficult for state policymakers to understand (Guthrie & Rothstein, 1999). One of the objectives of this chapter is to explain the intuition behind cost functions to help make them more accessible to policymakers. After all, complex statistical procedures are accepted in some policy arenas, such as revenue forecasting and program evaluation, and we see no reason why they could not become accepted in the design of state education aid formulas.

In the following sections we describe the use of cost functions and other methods to estimate cost differentials for resource prices, economies of size, and student needs.
Geographic Variation in Resource Prices

The impact of geographic variation in the prices of goods and services on the purchasing power of school districts has been recognized for decades (Brazer & Anderson, 1975; Chambers, 1978). Ten states incorporate geographic cost of education indices (GCEIs) into their school funding formulas (Lofgren, 2007), and the National Center for Education Statistics (NCES) has sponsored the development of GCEI for all school districts in the country using two different methods (Taylor & Fowler, 2006). Controlling for the compensation a district must offer to attract personnel of a given quality is particularly important for accurate cost estimation, because personnel compensation (salaries plus benefits) made up over half of current spending in the average school district in 2011 (Dixon, 2013, Table 6). In this section, we discuss the reasons for variation in resource prices and review the four most common approaches for estimating a GCEI. Each of these approaches attempts to measure the extent to which the cost of personnel varies across districts based on factors outside of districts’ control—not on variation in districts’ generosity. Variation in the price of inputs other than personnel has been largely ignored in the literature. This remains an important topic for future research.

Reasons for Geographic Variation in Resource Prices

The prices school districts must pay for resources can differ across school districts for several reasons: (1) cost of living; (2) labor market conditions; (3) local amenities; and (4) working conditions for employees. The higher the cost of living in an area, defined as the resources required to purchase a standard bundle of goods and services, the more school districts in the area must pay to attract employees of a given quality. Local labor market conditions can also affect the salaries districts are required to pay. If an area’s unemployment rate for professionals is high relative to the rest of a state, for example, then teachers and school administrators in that area may have relatively limited choices of alternative jobs and thus be more apt to accept school district offers with lower salaries and benefits. This type of situation may not persist in the long run, however, if teachers are mobile.

School employees, like other employees, may also be willing to sacrifice some compensation to have ready access, or proximity, to amenities including natural sites (coastline, lakes, mountains, or parks), transportation (highways, airports, or railway), cultural events, and other state or local public services. Finally, districts may trade off spending on factors related to working conditions against increased teacher compensation. Specifically, the salary required to attract instructional and administrative personnel may depend on the working conditions in the school district, which reflect both school policies (e.g., school size, class size, professional development spending, availability of instructional materials, school leadership, and culture) and factors outside district control (e.g., student characteristics such as socioeconomic background). Research has found that teacher mobility may negatively affect student achievement (Ronfeldt, Loeb, & Wyckoff, 2013) and may be influenced by working conditions (Boyd et al., 2011; Isenberg, 2010; Hanushek, Kain, & Rivkin, 2004; Ondrich, Isenberg, & Yinger, 2008).

Cost-of-Living (COL) Index

The cost-of-living (COL) approach estimates price differences for a “market-basket” of goods and services across geographic areas (Duncombe & Goldhaber, 2003). For each factor in the market basket, price data is collected by geographic area and a market basket is identified using data on consumer expenditure patterns. The final COL index is the spending required to purchase the market basket in each location relative to the state or national average. The use of a COL index as an education cost adjustment is based on the assumption that teachers compare real wages across districts, not nominal wages. This assumption implies that a high-COL district cannot attract the same quality teachers as a low-COL district without paying higher real wages.

The principal strengths of the cost-of-living approach are its conceptual simplicity and the fact that COL indices are based on private sector prices outside of district control (McMahon, 1996). This simplicity comes at a price, however. Even if a COL index accurately captures variation across locations in consumer prices and in
the wages required to attract teachers, school personnel do not necessarily shop or live where they work. In addition, COL indexes do not capture variation across districts in working conditions and local amenities, which can affect the compensation required to attract equal quality teachers. Moreover, COL data at the school district level are surprisingly difficult to obtain; existing national and state-level COL indexes provide no insight into within-state COL variation (Nelson, 1991; McMahon, 1996). As a result, some states, including Colorado, Florida, Oregon, and Wyoming, have developed their own regional COL indices and incorporated them into their school aid calculations (Bureau of Economic and Business Research, 2013; Corona Insights, 2012; Rothstein & Smith, 1997; Wyoming Department of Administration & Information, 1999).

**Competitive Wage Index (CWI)**

Another approach to estimating a GCEI is to use information on variation in private sector salaries (or full compensation), particularly for occupations similar to teaching, typically professional, managerial, or technical occupations. Some states, including Ohio, Massachusetts, New York, and Tennessee, have used measures of average private wages as cost adjustments in their education aid formulas. One disadvantage of this approach is that it assumes that private employees in these occupations are comparable on experience, education, and demographic factors across geographic areas.

A more appealing approach is to use detailed individual-level data on private employees to construct a private wage index that controls for employee characteristics. Taylor (2011) has applied this approach to Texas and Washington school districts. Combing data from the Occupational Employment Survey (OES) published by the U.S. Bureau of Labor Statistics and the Census, Taylor and Fowler (2006) developed a CWI for all school districts in the country by regressing salaries of college graduates on demographic characteristics (age, gender, ethnicity, education, and hours worked), occupational categories, and indicator variables for labor market areas. The regression results are used to "predict the wages that a nationally representative person would earn in each labor market area" (Taylor & Fowler, 2006, p. 9). The CWI is obtained by dividing the predicted wage by the state or national average wage. This CWI can be updated for other years by using the OES to estimate changes in wages across years by occupation and labor market area.

The comparable wage methodology is straightforward, and a carefully constructed CWI should capture the impact of cost of living, local amenities, and labor market conditions on the salary a district must pay to attract teachers of a given quality. The principal drawback to this methodology is that average private sector salaries are not likely to reflect differences in working conditions for teachers across districts. For example, private sector salaries in professional occupations are not likely to reflect the demographics of the student body in a district, the age and condition of school buildings, the extent of overcrowding in classrooms, and so on, which could be very important to the job choices of teachers.

**Hedonic Teacher Cost Index (TCI)**

A GCEI can also be estimated by separating the impact on teacher (or other employee) compensation of factors within and outside district control using statistical methods and then determining costs based only on external factors. What sets this approach apart from the others is that, to the extent possible with available data, it directly accounts for the effects of school working conditions on the salaries required to attract teachers to a district. It has been conducted for several states, including Alaska, Maryland, New York, and Tennessee (Chambers et al., 2004; Chambers, Taylor, & Robinson, 2003; Duncombe & Goldhaber, 2003). However, only Texas and Wyoming presently use a hedonic-based cost adjustment in their aid formulas (Taylor, 2010 and 2011).

The hedonic salary approach involves estimating a multiple regression model in which employee salary (or salary plus fringe benefits) is regressed on teacher characteristics, working conditions under district control (such as school or class sizes), and factors outside district control that are related to cost of living, labor market conditions, local amenities, and school working conditions. Characteristics of teachers typically include education, experience, gender, race, type of assignment, and certification status. Some studies include other measures associated with teacher quality, such as certification, test score performance, and ranking of the
college a teacher attended (Duncombe & Goldhaber, 2003). Amenity variables typically include distance to a central city and crime rates; working-conditions variables include district enrollment and student characteristics (e.g., disability, language proficiency, and poverty).

Hedonic models are typically estimated with individual teacher-level data using standard multiple regression methods. To construct a personnel cost index, the coefficients for discretionary factors are multiplied by the state average value for that factor, while coefficients for external (i.e., nondiscretionary) factors are multiplied by actual values for that district. The sum of these terms is the predicted salary required to attract an employee with average characteristics to a particular district. Because they have the most complete controls, hedonic salary models are likely to produce the most accurate estimate of the salary required to attract teachers with given characteristics to work in a district. However, even hedonic estimates face several difficult challenges (Goldhaber, Destler, & Player, 2010).

Perhaps the most difficult challenge is to fully control for teacher quality. Teacher characteristics included in existing studies capture several important dimensions of teacher quality, but these characteristics predict only a small share of variation in teacher quality as directly measured from teachers’ impacts on the test scores of their students (Harris & Sass, 2011; Rivkin, Hanushek, & Kain, 2005). Moreover, teacher quality is likely to be negatively correlated with concentrated student disadvantage, so imperfect controls for teacher quality will bias the coefficients of the student disadvantage variables toward zero. As a result, hedonic studies may systematically understate the impact of concentrated student disadvantage on the compensation a district must pay to attract teachers of a given quality.

In addition, actual teacher salaries may not correspond in all districts to the minimum salaries required to attract teachers with certain characteristics into the district. Some districts could be overly generous or particularly inept in bargaining, for example. Differences between actual salaries and minimum required salaries are signs of district inefficiency, and they could lead to biased results in hedonic salary models if the (unobserved) factors that lead to inefficiency are correlated with the explanatory variables in the model.

Another challenge is that readily available COL measures may reflect discretionary district decisions. For example, housing prices often account for most of the variation in private prices across geographic areas but they may partially reflect differences in perceived education quality across districts (Nguyen-Hoang & Yinger, 2011). Using Metropolitan Statistical Area (MSA) level housing prices reduces this endogeneity (Duncombe & Goldhaber, 2003). Some hedonic studies have used unimproved agricultural land as a COL measure to avoid the potential endogeneity of housing prices; however, agricultural land in central cities or inner ring suburbs often does not exist, and has to be imputed (Chambers et al., 2004). Private sector salaries can serve as a proxy for cost of living, labor market conditions, and some amenities, but are likely to be influenced by housing prices (and education quality) as well.

Several studies have attempted to address potential biases in hedonic salary models. Teacher fixed effects models have been estimated to control for unobserved teacher quality differences (Chambers et al., 2003). To account for the possibility of omitted compensation or working condition variables, some studies have included an estimate of the turnover rate in the model (Chambers et al., 2004; Duncombe & Goldhaber, 2003). Moreover, a few hedonic studies include variables to control for school district efficiency (Duncombe & Goldhaber, 2003). Finally, Boyd et al. (2013) develop a “two-sided matching model,” which separates school and teacher preferences and leads to the conclusion that “teachers prefer schools that are closer to home, have fewer poor students, and, for white teachers, have fewer minority students” (p. 106).

The TCIs calculated from hedonic salary models tend to display relatively little variation, because most of the variation in teacher salaries is explained by key factors in the models, usually education and experience, and because information on other determinants of teacher quality and on working conditions is incomplete. 

The limited impact of working conditions on hedonic TCI runs counter to recent research on teacher labor markets, which, as discussed earlier, finds that teacher mobility is influenced by the characteristics of the students they teach. More research is needed to resolve this apparent contradiction.

**Teacher Cost Indices from Cost Functions**

Resource prices, particularly teacher salaries, are key variables in education cost functions. The coefficient on the teacher salary variable indicates the increase in costs required to maintain student performance levels
when teacher salaries increase (holding other variables in the model constant). Using this coefficient and measures of teacher salaries by district it is possible to construct a teacher cost index (relative to the state average) that reflects variation in teacher salaries weighted by the impact of teacher salaries on spending.

Two different types of salary measures have been used in education cost functions: (1) private sector wage indices, such as a CWI (Reschovsky & Imazeki, 2001; Imazeki & Reschovsky, 2004a); and (2) actual teacher salaries for teachers with similar education and experience levels. Recognizing that teacher salaries can be set simultaneously with spending levels in the annual budget process, studies using actual teacher salaries often treat them as endogenous variables in estimating the cost function (Duncombe & Yinger, 2000, 2011a, and 2011b; Nguyen-Hoang & Yinger, 2014). In these studies, the teacher salary index may be based on actual salary or on salary predicted on the basis of factors outside a district’s control.

A teacher cost index derived from a cost function is similar in some respects to a CWI and should capture variation in employee compensation due to differences in cost of living, labor market conditions, and amenities across school districts. Because student characteristics are also included in the cost model, the teacher cost index is not likely to reflect the impact of working condition differences across school districts on the wages required to attract teachers. This impact may appear, however, in the estimated coefficients of these student characteristics; if so, it will appear when teacher cost indexes and pupil weights (discussed below) are combined.

The strength of this approach is that it produces a teacher cost index that both reflects variation in key factors affecting teacher salaries and is weighted by the impact of teacher salaries on spending. The accuracy of this approach depends, however, on the quality of the cost-model controls for student disadvantage and school district efficiency.

**Enrollment Size and Education Costs**

The 90-percent drop in the number of school districts in the United States since 1938 represents one of the most dramatic changes in education governance and management in the 20th century. While the pace of school district consolidation has slowed since the early 1970s, some states still provide financial incentives to encourage school district consolidation (Verstegen & Jordan, 2009) amid potentially strong opposition from local citizens (Weldon, 2012). However, operating aid formulas in a number of states compensate districts for small size or sparsity, thereby discouraging consolidation (Baker & Duncombe, 2004). In this section we briefly review the reasons for and the evidence on the relationship between costs and district size before discussing methods to estimate the cost effects of size.

**Reasons Costs May Vary with Enrollment**

Economies of scale are said to exist when the cost per unit decreases as the number of units increases. In education, the focus has been on economies of size, which refer to a decline in per-pupil expenditure with an increase in district enrollment, controlling for other cost factors. Several explanations have been offered for economies of size in education. First, some district services, such as central administration, are relatively fixed in the sense that the same central administrative staff may be able to serve a significant range of enrollment without a degradation of service. Economies of size might exist if larger school districts are able to employ more specialized labor, such as science or math teachers, which could improve the quality of instruction at no additional cost. Furthermore, teachers may be more productive in a large school district because they can draw on the experience of many colleagues. In addition, large districts may be able to negotiate relatively low prices for bulk purchases of supplies and equipment, or use their monopsony power to negotiate lower wages for their employees.

The existence of economies of size in education has been challenged for several reasons. First, some studies claim that the potential cost savings from consolidation are seldom realized because districts seldom lay off staff, salaries are often leveled up across the merging districts, and transportation costs actually increase (Hanley, 2007). Second, market concentration from consolidation may increase cost inefficiency, and thus
reduce potential cost savings (Gronberg et al., 2013). Third, large school districts tend to have large schools, which, according to some studies, lead to lower student performance (Kuziemko, 2006) by hurting staff morale, student motivation and involvement in school, and parental involvement (Cotton, 1996; Howley, 1994).

**Evidence on Economies of Size in Education**

A large literature on economies of size in education has emerged over the last four decades. Since this literature has been covered in depth in existing literature reviews (Andrews, Duncombe, & Yinger, 2002; Fox, 1981), we only summarize the main findings. The vast majority of evidence on economies of size has come from the estimation of education cost functions. The early evidence on economies of size found sizeable economies with the cost-minimizing size for an urban district with as many as 30,000 students (Fox, 1981). Recent cost-function research, which have addressed a number of methodological limitations with early studies, has also found that there may be sizeable economies of size in education, but that most of the cost savings from an increase in district enrollment are exhausted once enrollment levels of 2,000 to 4,000 pupils are reached. A few formal evaluations of the effects of school district consolidation on costs have been conducted. Zimmer, DeBoer, and Hirth (2009) find evidence of potential cost efficiencies from consolidations of school districts with enrollment below 2,000. Evaluating school district consolidations in New York from 1985 to 1997, Duncombe and Yinger (2007) find that doubling enrollment cuts operating costs per pupil by 61.7 percent for a 300-pupil district and by 49.6 percent for a 1,500-pupil district. Contrary to these studies, Gordon and Knight (2008) find that consolidation in Iowa "did not change pupil-teacher ratios, enrollments, or dropout rates," but they cannot control for student test scores.

**Methods for Estimating Cost Effects of Enrollment Size**

A common method used by states to construct scale adjustments is to estimate average district costs by enrollment class and then compare the average cost in a class to average costs in relatively large districts. Kansas used this strategy, for example, to develop "low enrollment weights," which are used in calculating operating aid (Duncombe & Johnston, 2004). The problem with this approach is that it does not consider factors other than enrollment size or sparsity, such as student performance, resource prices, topography, and student needs, each of which might affect spending differences across districts.

The cost-function method provides the most direct way to determine the relationship between enrollment and costs. By controlling for student performance, resource prices, student needs, and efficiency, cost functions have the potential for isolating the effects of enrollment size on cost differences. The key decisions in estimating economies of size in a cost function is selecting measures of student counts, and the functional form of the relationship between cost and enrollment.

Student counts used in aid formulas generally are of three types: (1) enrollment, which is the count of all students at one point in time (usually the fall); (2) average daily membership (ADM), which is an estimate of the average enrollment over the course of the year; and (3) average daily attendance, which measures the average number of students actually attending school. States adopt either one or a hybrid form of these three. In general, the difference between these student counts is quite small except in the large cities where attendance rates are often lower than enrollment.

The existence of economies of size implies a negative relationship between per-pupil spending and enrollment at least over some range of enrollment. However, it is likely that the rate of decline in per-pupil spending occurs more quickly at low-enrollment levels than at higher-enrollment levels because of relatively fixed costs, such as central administration. Several different functions have been used to account for the possible nonlinear relationship between enrollment and per-pupil cost. The most common approach is to use a quadratic function (the natural log of enrollment and its square) to model the relationship. Quadratic functions allow for the relationship between enrollment and per-pupil costs to go from negative to positive (that is, to be U-shaped), and cost-function studies have found diseconomies of scale as well as economies of scale (Reschovsky & Imazeki, 2001; Imazeki & Reschovsky, 2004b). In states with a few high-enrollment school districts (e.g., New York) the quadratic function can lead to estimates of large diseconomies of scale; some
studies have used cubic functions to reduce the effects of these large districts (Duncombe & Yinger, 2000). To allow for a more flexible relationship between enrollment and per-pupil costs, several cost-function studies have used enrollment “groupings” (such as 0 to 300, 301 to 2,000, etc.) instead of a quadratic function (Duncombe & Yinger, 2005). Flexible cost functions, such as translog functions, provide another alternative for specifying the enrollment-spending relationship by including both a quadratic term and a number of interaction terms between enrollment and other variables in the cost model (Gronberg et al., 2004).

Professional-judgment studies can also be used to estimate the effects of size on costs. (See Downes & Stiefel, this volume, for a more detailed discussion of this approach.) In professional-judgment studies, panels of education professionals are asked to estimate the resources required to produce a particular set of student performance results. Panels are typically asked to do estimates for prototypical schools or districts with different characteristics, such as enrollment size or poverty rates (Baker, 2005). The estimates for the prototypical districts can then be extrapolated to districts of different sizes to develop an economies-of-size estimate for all districts in a state. Using the results of professional-judgment studies in several states, Baker (2005) found that the shape of the per-pupil cost curve relative to enrollment was very similar to that found in cost-function studies.

**Student Disadvantage and Education Costs**

Extensive research on the determinants of student success in school indicates that peer characteristics, family composition, parental education and employment status, and neighborhood characteristics can significantly affect student success (Ginther & Pollak, 2004; Harris, 2010). In addition, student characteristics can affect the mobility decisions of teachers (Hanushek, Kain, & Rivkin, 2004), and hence both the quality of teachers and the costs of teacher recruitment and training. Moreover, districts with a high concentration of students living in poverty or with limited English proficiency face much greater challenges than other districts in helping their students reach academic proficiency. In this section, we discuss the types of student characteristics considered in the literature, the methods available for estimating the additional costs required to bring disadvantaged students to a given performance level, and how states have accounted for student disadvantage in their aid formulas.

**Measures of At-Risk Students**

The term “at-risk” implies that a student is at a higher risk of failing to meet educational objectives than are other students because of characteristics of the student or of his or her family or peers. The most widely used measure of “risk” or disadvantage is poverty. One key measure of poverty for education cost studies is the child poverty rate, defined as the share of a district’s school-age population (five to 17 years old) living in a poor household. The Census Bureau also provides intercensal estimates of child poverty. An alternative poverty measure more commonly used in education research is the share of students who qualify for a free or reduced-price school lunch as part of the National School Lunch Program administered by the U.S. Department of Agriculture. This measure has the advantage over the census poverty measure in that it is updated annually, but it is based in part on decisions by families to apply for participation in the program and on decisions by school districts to offer and promote this service. However, the percent of elementary students eligible for a free lunch is highly correlated with the child poverty rate, as reported in Duncombe and Yinger (2005) and Duncombe and Goldhaber (2003).

Other measures of student “risk” available in the decennial census include the share of children living with a single mother and the share of children living with a single mother who has an income below the poverty line and is not a high school graduate. States may also collect information on “Title 1” students (those eligible for Title 1 services). Students with limited English proficiency (LEP), also called English language learners (ELLs), may face significant challenges succeeding in school (see Rumberger & Gándara, this volume). Many states collect information on students who qualify for bilingual education programs or students that have been identified as needing language assistance. Unfortunately, however, there is no standard definition of LEP
across states, and the LEP data in some states are of questionable accuracy. An alternative measure is available from the Census, which collects information on the number of children ages five to 17 who live in households where English is spoken “not well” or “not at all” or of children who are living in households that are “linguistically isolated.”

One limitation of the data in most states is that they do not identify how many students fall into multiple at-risk categories. The data do not identify, for example, how many students are both from poor families and ELLs. Estimated cost differences across districts might be different if controls for these types of overlap were included. Moreover, the patterns of overlap might differ across states so that pupil weights estimated in one state do not apply elsewhere, even if the states’ performance measures are the same. More research on these overlaps is needed.

Students with disabilities or special needs generally require more resources than other students to reach the same student performance standards. To account for these extra costs, many states incorporate pupil weights or other adjustments for special needs students in their school aid formulas. In this volume, Parrish, Harr-Robins, and Chambers provide a detailed discussion of state policies, including aid formulas, to account for students with special needs.

Methods for Estimating Additional Costs to Educate Disadvantaged Students

Cost functions provide a direct way to estimate the impact of student disadvantage on the cost of education, holding student performance constant. To be specific, the coefficients on the variables measuring student disadvantages can be used to calculate pupil weights for each type of disadvantage (Duncombe & Yinger, 2005). A weight of 1.0 indicates for a given type of student disadvantage that it costs 100 percent more to bring a student in that category up to given performance standards than the cost for a student without disadvantage. Poverty weights estimated from cost functions vary within a state due to differences in methodology and across states because of differences in performance measures and other issues discussed earlier. One study (Duncombe & Yinger, 2005) estimates a range of poverty weights from 1.1 to 2.1 for New York, depending on the method; another (Duncombe et al., 2008) finds an average poverty weight of 0.55 for Kansas and 0.64 for Missouri. In the case of Missouri, but not Kansas, this weight is higher in central cities than in rural areas. In addition, estimated weights for students in poverty are between 0.23 and 0.31 for Texas (Gronberg et al., 2004) and between 0.3 (Imazeki, 2008) and 0.563 (Duncombe & Yinger, 2011b) for California.

Another approach for estimating the higher costs required to support at-risk students is to use professional-judgment panels. These panels can be asked to estimate the required resources needed to reach student performance standards for schools with different levels of poverty (or LEP shares). The differential in costs across these prototype schools can be used to develop rough estimates of pupil weights by student type. Pupil weights from professional-judgment panels are based on the judgments of professional educators and may be sensitive to the instructions given to the panels and to the ability of panel participants to identify the extra programs that would be required to bring at-risk students up to the specified performance standard (Rose, Sonstelie, & Richardson, 2004).

Baker (2006) compared several professional-judgment studies and cost-function studies for the same state and found that pupil weights produced from professional-judgment studies are generally lower than weights produced from cost-function studies. In Kansas and New York, for example, poverty weights calculated using the results from professional-judgment studies (Augenblick et al., 2002; Chambers et al., 2004) are half those calculated in cost-function studies (Duncombe & Yinger, 2005). One exception is the professional-judgment study in Maryland done by Augenblick and Myers (2001), which estimated pupil weights of 1.0 or higher for poverty and LEP.

How States Adjust for Student Disadvantage

Almost all state governments have some type of aid program that provides additional funds to districts with a relatively high concentration of at-risk students. Almost all states have a way to reimburse additional costs to educate students with special needs. Verstegen (2011) finds from her survey that 21 states incorporate the costs
of special education students into operating aid programs through the use of pupil weights, and 10 other states reimburse school districts (a certain percentage of) actual eligible costs incurred in providing educational services for these students. The weighted-pupil approach is also used to adjust the basic operating aid formula for poverty in 18 states and for students with limited English proficiency in 13 states (Verstegen, 2011). For poverty students, the lowest weight of 0.05 is used in Mississippi while the highest weight of 1.5938 appears in Georgia (Verstegen, 2011). Most states use eligibility for the National School Lunch Program as their measure of poverty, but some states use Census poverty estimates or federal welfare eligibility.

With the exception of Maryland, which recently implemented an aid formula with a poverty weight from a study that used the professional-judgment method, no state uses statistically estimated pupil weights in its formula. Nevertheless, pupil weights have been estimated for many states and have been considered in many policy debates.

The cost adjustments discussed in this chapter are designed to apply to a school district’s entire operating budget. Because they almost always represent a small share of a district’s budget, categorical aid programs for at-risk students are therefore unlikely, by themselves, to provide needy districts with the funds they need to meet an adequacy standard. Most states have not yet recognized that both incomplete cost adjustments in the operating aid formula and full cost adjustments applied to small categorical aid programs are incompatible with an adequacy objective.

Conclusions and Directions for Future Research

We find a broad consensus among scholars that the cost of achieving any given level of student performance is higher in some districts than in others because of (1) differences in the compensation needed to attract school personnel, (2) differences in enrollment size, and (3) differences in the concentration of disadvantaged students or those with special educational needs. We do not find a consensus, however, on the magnitude of these cost differences or on the best methods for estimating them. Instead, we observe an active literature with many different approaches to estimating costs and a lively debate about the strengths and weakness of each approach.

From our perspective, the core of this topic is the estimation of education cost models. Although scholars disagree about the details, these models are now widely used and have influenced the debate about state aid formulas in many states. The most difficult issues that arise in estimating these models are how to select performance measures and then how to control for the associated school-district efficiency. No consensus on the best approach has yet emerged and far more work on these topics is needed. Questions of variable selection and functional form also deserve more attention. Because they play such a critical role in state education aid formulas, pupil weights should continue to be a focus of this research.

A second focus of the literature has been on estimating teacher cost indexes. This topic also has important links to policy because a state aid formula cannot provide the resources needed to reach any student performance target without accounting for teacher costs. As we have shown, scholars have addressed this topic using a wide range of approaches with different strengths and weaknesses. The hedonic wage approach is the most appealing conceptually, but it also requires data that are often not available, and more research developing and comparing all the approaches would be valuable.

Finally, we are struck by both the clear link between cost estimation and the objectives of most state education aid formulas and the need for more work to make cost studies accessible to policymakers. Complex statistical procedures are accepted in some policy arenas, such as the revenue forecasts used in budgeting, but are only beginning to be accepted in the design of state education aid formulas. Because careful estimates of cost differentials can help policymakers achieve their educational objectives, we believe that further efforts to make these estimates accessible would be valuable.

Acknowledgments
Bill Duncombe passed away in May 2013, but he was a full partner in the previous version of this chapter and his contributions remain central to this version. We would like to thank Leanna Stiefel, Helen Ladd, and Ted Fiske for their excellent comments and suggestions. We are fully responsible for any errors and omissions.

Notes

1. Published cost function studies have been conducted for Arizona (Downes & Pogue, 1994), California (Duncombe & Yinger, 2011b; Imazeki, 2008), Indiana (Zimmer et al., 2009), Kansas (Chakraborty & Poggio, 2008; Duncombe et al., 2008), Kentucky and Maine (Lee, 2010), Massachusetts (Nguyen-Hoang & Yinger, 2014), Missouri (Baker, 2011; Duncombe et al., 2008; Duncombe & Yinger, 2011a), New York (Duncombe & Yinger, 2000 and 2005), Texas (Gronberg, Jansen, & Taylor, 2011; Imazeki & Reschovsky, 2003, 2004a, and 2006); and Wisconsin (Imazeki & Reschovsky, 2003; Reschovsky & Imazeki, 2001). A detailed review of education-cost-function studies is provided by Gelebiewski (2011).

2. More formally, equation (1) needs to be estimated with two-stage least squares regression, which requires "instruments." These instruments are variables that influence $E$ but do not influence $S$ directly. Recent studies use instruments that measure the determinants of the demand for $S$, such as socioeconomic characteristics, in comparable school districts, which form a point of comparison for voters and school officials (Duncombe et al., 2008; Duncombe & Yinger, 2005, 2011a, and 2011b). Several studies, for example, Imazeki and Reschovsky (2004a) and Reschovsky and Imazeki (2003), use a district’s own income and tax price as instruments. These instruments are not legitimate, however, because, as shown below, they are determinants of efficiency and therefore influence $E$ directly. As in Nguyen-Hoang (2012), and Ross and Nguyen-Hoang (2013), an annual performance rank index, which is derived by dividing districts’ performance into equal percentile-based thirds and assigning the thirds ranks of one, two, and three could be potential instruments for performance. Overall, the choice of instruments is an important topic for future research.

3. The main cost frontier method to estimate efficiency is data envelopment analysis (DEA) as in Chakraborty and Poggio (2008), Flavin et al. (2012), and Ruggiero (1998, 2001, 2007). This DEA approach is not without weaknesses: inability (as a result of no standard errors) to test hypotheses, and potential bias from measurement error and model misspecification (Smith, 1997). Another approach is a stochastic frontier regression (SFR) (Gronberg et al., 2011). Ondrich and Ruggiero (2001) show, however, that SFR and ordinary least squares (OLS) yield the same results except for the intercept. In other words, SFR is appropriate only if efficiency is uncorrelated with any explanatory variables.

4. Ruggiero (1998) shows how to separate efficiency and cost measures in DEA, but his method requires more extensive data than are usually available.

5. In a cost-function context, it is not possible to separate inefficiency associated with "wasteful" spending from inefficiency associated with spending on performance measures other than those included in $S$. It follows that a given school district could be deemed inefficient in providing some types of student performance, say math and English proficiency, and efficient in providing others, say art and music literacy.

6. Most studies use a variant of the Cobb-Douglas function, which is multiplicative in form. The Cobb-Douglas function assumes that the elasticity of substitution between all inputs is equal to one, and that the elasticity for economies of scale is constant at all levels of output.

7. One of the most popular flexible cost functions used in empirical research is the translog cost function. A translog cost model includes squared terms for each input price and outcome, and adds interaction terms between all factor prices and outcomes. Gronberg et al. (2004) also include a number of interaction terms between outcomes, teacher salaries, and nonschool factors. In all, they have over 100 variables in their cost function for Texas compared to 18 variables in the Texas cost model estimated by Imazeki and Reschovsky (2004a).

8. Downes (2004) argues that rejecting the cost-function method because it is not easy to understand "means that other methodologies should be used in place of the cost-function methodology, even if the cost-function methodology is theoretically sound and is most likely to generate valid estimates of the spending levels needed to meet the standard. Taken to the extreme, this argument implies that, in choosing a method to determine adequate spending levels, one is better off choosing a method that is easy to understand but wrong than a method that is difficult to explain but produces the right answers" (p. 8).

9. Colorado has recognized this possibility by calculating cost of living for "labor pool areas," which are designed to reflect where teachers in the district live, rather than where they work.

10. Some hedonic salary studies have not included any measures of student characteristics (Chambers et al., 2004), and a number of studies do not include measures of student poverty (Chambers et al., 2003).

11. Specifically, the cost models are estimated with two-stage least squares regression, and "instruments" are identified for teacher salaries. Instruments have included private sector salaries, county population density, and teacher wages in surrounding or similar districts.
 Economies of scale can occur either because of increasing returns to scale, which arise when a 1 percent increase in inputs leads to a more than 1 percent increase in output, or because prices for certain resources used by the school district decline with the amount of resources used.

Children with incomes at or below 130 percent of the federal poverty line are eligible for a free lunch, and students between 130 and 185 percent of the poverty line are eligible for a reduced price lunch. In addition, households receiving food stamps or Temporary Assistance to Needy Families (TANF) are also eligible for a free lunch. A description of the program and eligibility requirements is available at fns.usda.gov/nslp/national-school-lunch-program.

A recent study by Dahl and Scholz (2011) found that in the 2002–03 school year, only 72 percent of eligible children nationwide received free or reduced-price lunches while about 12 percent of ineligible children received free or reduced-price lunches this school year.

Skin color is not literally a cost factor, of course, but in our society skin color is often highly correlated with student disadvantages. In this context, Baker (2011) provides a thoughtful discussion of the use of race in a cost-function regression.

Language ability is estimated from the long form of the census, in which individuals are asked if they speak a language other than English and are asked their ability to speak English. A household in which all members of the household 14 years or older do not speak English well and speak at least one other language than English are classified as linguistically isolated.

These weights are our estimates based on the estimated range in marginal effects for free lunch students in Table 3 in Gronberg et al. (2004) divided by average spending per pupil. They should not be attributed directly to the authors of that study. These relatively low weights may be due in part to the fact that this study interacts the child poverty rate with several other variables. Several of these interaction terms are not statistically significant. An alternative way to separate cost and efficiency is developed by Ruggiero (1998 and 2001) based on the two-step methods developed by McCarty and Yaisawarng (1993). In the first stage, Ruggiero compares the spending per pupil of each district to a cost frontier (districts with equivalent outcomes and lower spending) using data envelopment analysis (DEA) discussed earlier in endnote 3. The index produced from the first-stage DEA captures both inefficiency and cost differences across districts. Ruggiero then regresses this index on a set of cost factors, and the predicted value from this regression is his estimate of cost of education index.

Yinger and Duncombe (2004) developed poverty weights for New York using information in the professional-judgment study (Chambers et al., 2004); these estimates, which should not be attributed to the authors of the study, range from 0.37 in middle school to 0.81 in elementary school.

References


Introduction

Public education in the United States is big business. Current expenditures for public K–12 education are estimated to total $619 billion in 2014–15 (U.S. Department of Education, 2013). These revenues are raised as part of the larger federal fiscal system in the United States where governments at the local (such as cities and school districts), state, and federal levels all raise and spend public tax dollars.

Although much of the responsibility for providing elementary and secondary education services rests with the nearly 14,000 local school districts across the nation, almost half the financing for K–12 education is now provided by the governments of the 50 states. Only since the late 1970s has the state become an equal partner with local school districts in financing education. The federal role in education funding was negligible until passage of the Elementary and Secondary Education Act in the mid-1960s. Since that time between 7 and 10 percent of K–12 education funding has come from the federal government.

States are now playing a more significant role in the finances of schools for a number of reasons. In response to the first wave of school finance lawsuits starting in the 1970s (see Koski & Hahnel, this volume), states have used state revenues to equalize differences in the revenue raising capacity of their local school districts. States have also stepped in to provide local property tax relief. Often these two efforts work hand-in-hand, with increases in state revenues partially offsetting local property taxes and partially equalizing educational spending. With subsequent rounds of court cases, states have also expanded their role by trying to ensure adequate spending and to promote efficiency and productivity.

School districts raise most of their educational revenues through local property taxes. Because property is fixed in location, and values tend to change slowly, such taxes offer relatively small units of government—such as school districts—a stable source of revenue (Monk & Brent, 1997). States, which have a larger base upon which taxes can be levied, typically rely on other taxes, especially sales and income taxes, to finance their operations. These broad-based statewide taxes enable states to distribute funds to offset local disparities in wealth and in student needs. Thus, there has been a shift from local sources to state and, to a lesser extent, federal revenue sources over the last 40 years. Simultaneously, the emphasis of school finance funding formulas has changed from concern solely about equity to consideration of adequacy as well.

Equity

School finance has long been concerned with the equitable distribution of funds to school districts. The traditional problem facing states in funding their schools is the diversity in the size, financial capacity, and student needs of the school districts within the state’s boundaries. General definitions of equity include equal revenue per pupil (horizontal equity), additional spending per pupil for those with identified special needs, such as students from a poverty background, students with disabilities, and English learners (vertical equity), and a lack of connection between revenues per pupil and measures of local fiscal capacity (fiscal neutrality).

The most obvious challenge to the equitable distribution of funds to schools is the frequently large disparity in the ability of school districts to collect property taxes. Property taxes are typically assessed at a uniform rate within a taxing jurisdiction on all real property located within that jurisdiction’s boundaries. The amount of money that can be raised for each pupil varies substantially among districts within any state.
In the late 1960s this variation in revenue per pupil was challenged under the equal protection clauses of the U.S. Constitution and state constitutions. Two legal challenges—McInnis v. Shapiro (1968) in Illinois and Burruss v. Wilkerson (1969, 1970) in Virginia—failed to establish a standard to assess and measure educational need. In 1970, Coons, Clune, and Sugarman proposed treating the use of property wealth per pupil as a suspect classification as a way to satisfy judicial scrutiny. This strategy was first used in Serrano v. Priest in California in 1971.

The U.S. Supreme Court ruling in San Antonio v. Rodriguez in 1973 eliminated the use of the Equal Protection Clause of the U.S. Constitution when the Court declined to apply the strict scrutiny test. The Court determined that education was not a fundamental right under the U.S. Constitution and that low property wealth was not a suspect class.

The focus of subsequent school finance litigation shifted to the states. Following the logic established in Robinson v. Cahill in New Jersey in 1976, state courts focused on the education clauses of their state constitutions. While the individual state clauses vary considerably (McUsic, 1991; Underwood, 1995), many state courts have used terms such as "thorough and efficient," "thorough and uniform," or "general and uniform" as the standard to which school finding systems must be held. Regardless of whether the courts ruled that systems were unconstitutional, most states moved to increase the equity of their school funding systems, generally through increases in overall education funding (see Corcoran & Evans, this volume).

**Adequacy**

As spending for K–12 education increased in the 1980s and 1990s, policymakers and academics called for more effective use of the funds and better understanding of how dollars could lead to improvements in student learning. At the same time, the national discourse shifted from educational equity to educational excellence, and with it a focus on academic standards and student outcomes. Questions were raised about whether existing levels of funding were adequate to ensure that all students could meet more demanding educational standards. In contrast to equity, which focuses on the relative distribution of funds, adequacy is generally thought of as providing a sufficient level of resources to provide a set of educational programs and strategies so that each child has an equal opportunity to meet the state’s educational performance standards.

Beginning with court rulings in West Virginia (Pauley v. Kelly, 1979), Washington (Seattle Sch. Dist. 1978), and, most importantly, Kentucky (Rose v. Council for Better Education, 1989), the adequacy movement took hold as the most common approach for challenging the school funding systems in the states. Interest in adequacy has been encouraged by the growing number of successful law suits challenging the adequacy of state school funding systems; by state efforts to meet the requirements of the federal No Child Left Behind Act, which requires all students meet state proficiency standards by 2014; and by state adoption of the Common Core State Standards.

There is a growing body of literature on how to determine what an adequate level of spending on education should be (see, e.g., Downes & Stiefel, this volume; Odden & Picus, 2014). As described by Downes and Stiefel, four approaches to estimating adequacy have been developed to date—each with strengths and weaknesses. Over the years, the distinction between at least three of the four methods—evidence-based, successful districts/schools, and professional judgment—have diminished, as analysts have sought to show how the three are connected and seek to link elements in the funding formula to strategies schools use to boost student learning (Picus et al., 2013).

**School Finance Formulas**

State school finance formulas link states and school districts in a fiscal partnership for funding public schools. These funding systems serve two major purposes: (1) to compensate for wealth-based disparities in educational revenues; and (2) to address cost differences based on student and district needs. In this section we first describe three general forms of school finance formulas: foundation programs, guaranteed tax bases, and combination programs. We then discuss issues of cost differences and consider approaches that states use to
meet these needs.

Since most school districts raise their own source revenue from property taxes, state aid formulas are primarily designed to provide more state aid—for the same level of effort—to districts with low property wealth per pupil and less state aid to districts with higher property wealth per pupil. The fundamental goal of these formulas is to provide equal access to total revenues per pupil for school districts across a state. In some instances the goal is to ensure a minimum level of per-pupil revenue (achieved through a foundation program), while in other instances the goal is to provide equal resources for equal effort, leaving the choice of total per-pupil spending to the local school district (achieved through a guaranteed tax base or GTB program). Combination programs join elements of foundation and GTB programs to equalize school district per-pupil revenues by ensuring a minimum revenue level in each district and providing equal access—at least up to some agreed-upon level—to additional resources should local taxpayers want to make additional efforts on their own.

**Foundation Programs**

In general, a foundation program establishes a minimum per-pupil revenue level at a set property tax rate and then guarantees that districts levying that tax rate will receive the foundation level of revenue. State aid is the difference between the revenue guaranteed as a “foundation” and the amount raised locally by the established tax rate. The design of foundation programs raises several policy issues. The major issue is establishing the foundation level itself. In the past, this level has often been decided politically, based on the state revenue available for education in the budget year. The recent advances in school finance adequacy, however, provide methods for estimating what the foundation level should be. As we discuss below, once a foundation level is established, adjustments may be made for differences in student needs and to compensate for differences in the costs of educational services and materials that are beyond the control of a school district.

In addition to the foundation level of funding, state policymakers need to agree on the required tax rate expected of local school districts, whether or not local school districts can exceed the revenue level established through the foundation program, what to do if a local school district’s property tax revenues exceed the foundation guarantee at the minimum required tax rate, and whether or not to equalize revenues in excess of the foundation guarantee if districts elect to raise additional funds. This latter issue is discussed under combination programs below. In some states the “excess” revenues are subject to recapture—or redistribution to other school districts—while in other states, districts are allowed to keep all the additional funds they raise. In Texas, where recapture is not an option under the state’s constitution, the state devised a structure that requires wealthy districts either to transfer the value of part of their tax base to another school district for taxation purposes or to make payments to other districts or to the state when property tax revenues exceed the foundation guarantee.

In simple terms, a foundation program provides aid to each school district in inverse relationship to its property wealth per pupil, as represented in the following equation:

\[ SAPP = FLPP - (RTR \times PVPP) \]

where

- **SAPP** = state aid per pupil
- **FLPP** = the foundation level of revenue per pupil
- **RTR** = the tax rate each district is required to levy
- **PVPP** = the property value per pupil of the school district

Thus, the amount of state aid given to a particular district is inversely proportional to the value of the district’s per-pupil property value. Under a foundation model, the foundation level (FLPP) and the required tax rate (RTR) are state policy decisions. Local districts can choose to spend more than the foundation level by taxing their own property value at a tax rate exceeding the required tax rate.

Arkansas provides an example of the use of such a formula. Based on a 2003 adequacy study, the Arkansas Legislature determined that the cost of funding an adequate education program amounted to $5,400 per pupil
(Odden, Picus, & Fermanich, 2003; Odden, Picus, & Goetz, 2006). This figure was then funded through a foundation program with a required property tax levy of 25 mills in each school district (a mill is a dollar of tax per $1,000 assessed property value). To this foundation amount for the base program, Arkansas added two major new categorical programs: $480 for each student eligible for free and reduced price lunch and $195 for each English language learning student. Funds for students with mild and moderate disabilities were included in the calculation of the base foundation level.

More recently, North Dakota took a similar approach. It began with an adequacy study completed in 2008 (Odden et al., 2008) that found that an adequate funding level was $7,293 per student. The state enacted portions of the overall study’s recommendations in the next two biennia. The 2013 legislature completed the transformation of the school finance system by restructuring the state’s school finance system into a foundation program using $8,810 per pupil as the foundation level, funded with a 60 mill required local property tax rate (applied to a tax base that is approximately 5 percent of market value). Weights for special pupil needs including at-risk students, students with disabilities, and students learning English provide extra resources for the additional services those students need. To ensure that the system continues to provide an adequate funding level, the state launched a recalibration study in 2014, the results of which were not finalized by the time this chapter was written.

Foundation programs, which are used in 45 states (Verstegen, 2011), have several attractive features. Foundation formulas are unique in having a base program guarantee as a critical variable. They began as programs designed to provide a minimum quality education program, but today they can be used to guarantee a higher-quality program. For example, states could set the foundation level at the adequate base spending level identified through an adequacy study. Foundation programs are also fiscal-capacity equalizing in that they provide state aid in an inverse relationship to local property value per pupil, addressing the key structural problem of school finance—the disparity across districts in the local property tax base. Their key defect—at least from an equity perspective—is that they may allow local spending above the foundation program. If the base program is low, or does not keep pace with rising costs or educational needs, these local fiscal add-ons—financed entirely with local property tax revenues—restore some of the linkage between property wealth and education spending.

**Guaranteed Tax Base Programs**

Although foundation programs guarantee a fixed level of funding, guaranteed tax base (GTB) programs, which are very similar to district power equalizing programs, ensure a fixed level of revenue per pupil for a given tax rate. For example, if the guaranteed tax base is $600,000 of assessed value per pupil, a ten mill levy will raise $6,000 per pupil. A GTB program guarantees that any district whose wealth is less than $600,000 per pupil would receive state aid to make up the difference between $6,000 and what is raised through its property taxes. Mathematically this can be expressed as follows:

\[
SAPP = DTR \times (GTBPP - PVPP)
\]

where

- \(SAPP\) = state aid per pupil (as above)
- \(DTR\) = the district tax rate (note this is different from a required tax rate used in the foundation program in that it is chosen by the district)
- \(GTBPP\) = the property value per pupil level guaranteed to each district or the guaranteed tax base per pupil
- \(PVPP\) = property value per pupil

Another way to look at this is to express total district revenue per pupil (TRPP) as follows:

\[
TRPP = (DTR \times PVPP) + (DTR \times (GTBPP - PVPP))
\]

which can be simplified to be:

\[
TRPP = DTR \times GTBPP
\]
Under the foundation program described above, the state sets the foundation spending level (FLPP) and the required tax rate (RTR) districts must levy to receive that funding, while under a GTB program, the state establishes the level of property value it will guarantee (GTBPP) and lets the district establish its own tax rate (DTR). The education spending preferences of each district dictate the tax rate and level of spending. All districts with property wealth per pupil below the GTBPP receive state aid.

Because a GTB program allows different local decisions on education per-pupil spending levels, equality of spending is not its focus. Without a requirement for a minimum school tax rate, GTB programs do not even require a minimum education expenditure per-pupil level. Still, in most situations where GTB programs have been enacted, they have increased expenditures in all but the school districts with the lowest tax rates. However, it should be emphasized that a GTB program is incompatible with the standard horizontal equity principal for students that requires equal spending per child. Instead it is a funding approach that offers a version of fiscal neutrality.

Initially implemented in a number of states in the 1970s to remedy the major source of school spending differences across states—unequal access to the local tax base—GTB programs were a highly visible policy alternative. However, as Feldstein (1975) showed, GTB programs do not lead to the same level of spending in all districts due to a price effect (i.e., the variation in the “cost” to local taxpayers of providing one more dollar of spending to education will vary depending on property wealth of the district). GTB programs also appear to have been successful in stimulating more education spending, but produced what Odden and Picus (2014) term the “new” school finance equity issue. In the 1970s, property poor school districts usually had low spending per pupil even though they had high tax rates, while property rich districts usually had high spending per pupil with low tax rates. The GTB was expected to help poor districts by allowing them to lower their tax rates and simultaneously increase their spending. But in some states with GTB programs (i.e., Illinois, Missouri, and Wisconsin), spending differences, though attenuated, continued to exist and the patterns changed. Today in those states, property rich districts have high spending levels and high tax rates; average wealth districts tend to have average spending levels and average tax rates; and below average wealth districts tend to have below average spending and below average tax rates. Spending differences still exist, but they are driven by differences in local tax rates for schools. An issue for these states is whether such spending differences are a problem. Even though they were not caused primarily by differences in the local property tax base for schools, they are likely related to overall school district fiscal capacity.

Combination Formulas

The discussion above suggests that efforts to ensure a minimum (or adequate) level of revenue to school districts are best served with foundation programs, while efforts to reduce the role of the local tax base in causing spending differences are best served though GTB formulas. As their name suggests, combination programs use the strengths of each approach to design a system that ensures a minimum level of resources for all children, but provides an equitable way for any district that chooses to do so to increase its spending above the base foundation level. The foundation portion of the combined program first ensures a base spending level, usually above what had been a minimum level. This base spending level remedies a possible shortcoming of pure GTB programs that do not require a minimum spending level. The GTB portion of the combined program ensures equal education spending per pupil for equal tax rates above the foundation-required tax rate. This component remedies a defect of a minimum foundation program: unequalized spending above the foundation base.

Missouri has had a two-tiered, combination foundation and guaranteed tax base program since the late 1970s. When established, the Missouri formula set the foundation revenue level at just below the previous year’s statewide average expenditure per pupil. For the second tier, the legislature used a GTB program, setting the wealth guarantee at that of the district at the 95th percentile of property wealth. Over time wealthy districts increased their tax rates to fund higher education expenditures, while low-wealth districts—despite the relatively lower local cost—did not. Consequently, funding the GTBPP at the 95th percentile of property wealth did not overburden the state system. Spending differences between the wealthy and poor districts, however, increased because of different local preferences for education and willingness to pay.

The combination approach was used for other new school finance formulas established during the early
1990s. Both Texas and Kentucky, under court order to revise their school finance structures, enacted combination foundation and guaranteed tax base programs. In Texas, the 1989–90 foundation program provided a base spending or foundation level equal to about 42 percent of the statewide average expenditure per pupil. The guaranteed tax base program was set just below the statewide average per-pupil property value. Texas placed a tax rate cap on the GTB component of the formula, providing GTB aid for just an extra 3.6 mills above the foundation required tax rate. Districts were also allowed to levy higher tax rates, for which revenues were derived solely from the local tax base up to a maximum of $1.50 per $100 of assessed value or 15 mills. It is important to note that as this is written, the Texas school funding system is being challenged again in court. While the case revolved around adequate funding levels, the equity of funding was also part of the trial record. It is possible the Texas funding formula will change in the near future.

Kentucky enacted a similar but three-tiered combination program. The 1989–90 foundation base was set at about 77 percent of the statewide average per-pupil spending. Kentucky also put a GTB on top of the foundation program, setting the wealth guarantee at about 150 percent of the statewide average. This GTB program, however, included two tiers, each with its own type of tax rate cap. The first tier limited the additional tax rate beyond which districts could not receive GTB aid by giving school boards the flexibility to increase spending (and thus the local tax rate) by 15 percent over the foundation base and still receive GTB aid. In addition, taxpayers could increase spending by a local vote (and thus the local tax rate) by another 30 percent of the total amount generated by the foundation level and the first tier, but they would not be eligible for GTB aid for this second 30 percent spending boost. Thus, expenditures above the foundation base are limited to an additional 49.5 percent, 30 percent of which is fiscal-capacity equalized by a GTB.

As an example of how this works, suppose the base foundation level were set at $4,000 per pupil. A district would receive this level of funding once it made the required tax effort. It would be allowed to increase its revenue to $4,600 per pupil (a 15 percent increase) and have that $600 equalized as well. If the district elected to spend beyond this level, it could increase total per pupil revenue by an additional 30 percent or an additional $1,380 per-pupil (30 percent of 4,600 is 1,380). This would bring the total available revenue to $5,980 per pupil, before adjustments for student characteristics, nongeneral fund resources, and other categorical resources from the state and federal governments.

A combination foundation and GTB program offers many advantages. Both components of the program require local matching funds and provide for fiscal capacity equalization. A base spending level is guaranteed, providing some element of expenditure equity. The ability to spend above the base is possible on an equal basis for rich and poor districts alike, thus providing the potential for a fiscally neutral system depending on the decisions made by local school districts (local choice). It will not break the relationship between wealth and expenditures, however, if wealthy districts continue to choose higher tax rates and thus higher spending. While a combination program is fiscally neutral in its design (ex ante), it may not be in its operation (ex post) due to local preferences and willingness to pay. If a state enacts a cap on the level of extra revenues, such as the 30-percent cap in Kentucky, the program might be more appealing to those who champion horizontal equity for children.

The second-tier GTB has in some instances functioned as an incentive to spend more, and primarily by suburban districts with above average wealth. As a result, the two-tier system, just like an unbridled GTB program, can create a system that generally results in a combination of low-spending, low-tax-rate, low-wealth districts and low-spending, high-tax-rate and high (or above-average) wealth districts. Some would not consider this to be a fair system.

The Cost Dimension

To this point the discussion has focused on equalizing access to dollars under the assumption that all students and school districts are alike in terms of needs. This, of course, is not the case, and school funding systems generally adjust for differences in student characteristics and district characteristics—two issues over which the district has little control.

There is considerable evidence that students from different backgrounds require different levels of educational services. Today, state school finance formulas provide additional resources for students from low-income households, for children who are English language learners, and for children with disabilities that
impact their ability to learn and/or to participate in a regular school program.

There are two basic approaches for adjusting aid formulas for student characteristics. One method provides additional aid in the form of categorical grants or aid intended to be used for the purpose of providing services for children with specific needs. Examples of such programs include funding for special education for children with disabilities and compensatory aid programs for children from low-income families. In both instances, districts receive additional dollars based on the number of children (and sometimes on the intensity of the needs of those children) to fund programs that meet the specific needs of those children.

A second approach is to “weight” children based on their identified needs. Thus in a state aid formula, children from low-income families might receive a weight of 1.25 and limited English proficient children a weight of 1.10, or “extra” weights of 0.25 and 0.10 respectively. When the enrollment of a district is determined, the total number of weighted students receives state funding. Details on incidence, needs, and costs of these programs are provided in Section VII of this volume, and determination of the appropriate weights is a complex undertaking. Some states rely on relatively simple weighting schemes, offering one weight for low-income children, another for English language learners, and a set of weights for special education based on the disabilities of children. Other states have established more complex weighting formulas, establishing varying weights for different grade levels, vocational education programs, or even variable rates for the concentration of low-income children in a school or district.

A number of district characteristics that affect the cost of education are beyond the control of local school authorities. Perhaps the most debated are regional price differences across a state and the types of adjustments to make (see Duncombe, Nguyen-Hoang, & Yinger, this volume). Adjustments in school finance formulas for price differences across geographic areas within a state are fraught with political considerations. While it makes sense to compensate districts in high-cost areas for those additional costs, it is often difficult for policymakers to accept the reality that one dollar in one part of the state has less buying power than one dollar in another part of the state.

Measuring Equity in the States: Longitudinal Analyses of Equity

How have school finance systems done over time in terms of equity objectives? A considerable body of literature addresses this question. Several studies have analyzed the status of school finance equity across the 50 states over the last 30 years (see, for example, Corcoran & Evans, this volume; Schwartz & Moskowitz, 1988; Stiefel & Berne, 1981). The early studies had conflicting results, depending on the time frame studied and on the equity object selected and statistic used. A more recent study that covers the period 1972 to 2002 found that within state fiscal disparities had been reduced over this period by 15–19 percent in states with court-ordered reform, relative to those without such mandates (Corcoran & Evans, this volume). The examples presented below provide a more detailed look at the impact of school funding formulas enacted in response to court cases in Kentucky, New Jersey, and Vermont.

**Kentucky**

As described in the preceding section, Kentucky’s school finance system (SEEK) combines foundation and GTB formulas. Picus et al. (2004) found that the equity of Kentucky’s funding system steadily improved between 1990–91 and 1999–2000 and that the state had achieved high levels of horizontal and vertical equity and fiscal neutrality. For example, the coefficient of variation (CV) for revenues per pupil decreased from 0.143 to 0.108. Similarly, the CV for revenues per weighted pupil (reflecting special student needs) dropped from 0.136 to 0.104. Analysis of the McLoone index, a measure of distribution at the low end of the revenue spectrum, shows a high degree of equity among those districts, rising from 0.94 to 0.96 for revenues per weighted pupil. Although the formula allows for up to a 50 percent difference in revenues per pupil for those districts that add both Tier I and Tier II revenues, the restricted range was less than $1,500 per pupil. Finally, the link between property wealth and revenue per pupil was nearly gone by 2000. The correlation between these two variables dropped more than 50 percent from 0.494 to 0.220. The elasticity, which measures the strength of this
relationship, also decreased considerably from 0.15 to 0.07.

This trend toward improved equity did not continue. An analysis of equity for the 2012–13 school year found the Federal Range Ratio increased to 0.491 and the coefficient of variation was back up to 0.143. The Gini coefficient increased to 0.079, and the McLoone index fell to 0.909. In terms of fiscal neutrality, the 2012–13 per-pupil revenue data showed a correlation of 0.743 and an elasticity of 0.300, suggesting that the relationship between property wealth and per-pupil revenues had grown stronger in the last decade.

Thus, while SEEK initially achieved the equity goals set out by the legislature in 1990, there was deterioration in measures of equity between 2000 and 2013. The question remains, however, whether Kentucky districts have adequate resources to allow students to meet state performance standards. Adequacy studies have suggested that the state would need to increase funding for education from between $740 million to as much as $2.1 billion (Picus, Odden, & Fermanich, 2004; Odden, Fermanich, & Picus, 2003; Verstegen, 2003). A new adequacy study is being conducted in Kentucky as this volume goes to press.

**New Jersey**

New Jersey represents a particularly interesting case. In nearly two dozen decisions handed down over 35 years, the New Jersey Supreme Court has sought to ensure that all students in New Jersey, particularly in distressed urban areas, have equal access to a quality education. The decisions in Robinson v. Cahill and the initial decisions in Abbott v. Burke focused on equalizing educational spending, particularly between the state’s poor urban and wealthy suburban school districts (Goertz & Edwards, 1999). In the absence of other measures, the court used spending in the state’s highest-wealth districts as its benchmark for the level of resources needed to ensure equal educational opportunity in poor urban districts. In 1990 it ordered the legislature to design a funding system that would: (1) equalize spending for the regular education program between the state’s poorer urban districts (Abbott districts) and property-rich districts; and (2) provide additional funds to meet the special educational needs of the urban districts in order to redress their disadvantages (Abbott v. Burke, 1990, 1994, 1997). The court did not address spending disparities in poor rural or in middle-class districts in the Abbott case.

In 1984–85, when the first Abbott decision was handed down, the Abbott districts spent a few hundred dollars per pupil more than the poor, non-Abbott districts, but $800 dollars per pupil less than the middle wealth districts, and over $2,000 less than the high-wealth communities. The school tax rates in the Abbott districts were also nearly 50 percent higher than those in the high-wealth districts. By 2007–08, the Abbott districts were spending about $1,300 per pupil more than the high-wealth districts, and considerably more than both the poor non-Abbott ($4,000) and middle wealth ($3,000) districts, while levying much lower tax rates. Thus, while parity had been achieved between the Abbott and high-wealth districts, as intended by the court, both spending and taxpayer disparities remained among the non-Abbott districts (Goertz & Weiss, 2009). Although changes in New Jersey’s school finance system had done little to alleviate wealth-based disparities in education spending among the non-Abbott districts, they had made the distribution of education spending somewhat more equitable statewide. Between 1984–85 and 2005–06, the coefficient of variation dropped from 0.183 to 0.165, the federal range ratio dropped from 0.816 to 0.655, and the McLoone index rose from 0.859 to 0.887 (Goertz & Weiss, 2007). The coefficient of variation and federal range ratio remained relatively high, however.

In 2008, the School Finance Reform Act (SFRA) replaced the Abbott remedies with one formula applicable to all districts in New Jersey. The new formula, which was declared constitutional in 2009, eliminated any special funding for the Abbott districts (except for facilities aid) and, by increasing the weights for at-risk and LEP students, provided significantly more aid to the poor and middle-wealth non-Abbott districts. This new targeting of aid, combined with substantial underfunding of the formula, closed expenditure disparities between low- and middle-wealth districts, but reopened the gap with the wealthiest communities (Education Law Center, 2014).

**Vermont**
Vermont offers another example of a state that has improved equity over time in response to a lawsuit. In 1996 the state’s supreme court held the Vermont funding system to be unconstitutional. Over a period several years, the state enacted Acts 60 and 68 implementing a funding system that equalized school district revenues, provided income sensitive caps on local education taxes, and enabled local districts to increase revenues through an equalized system. In 2012, Picus et al. concluded that the Vermont school funding system had achieved a high degree of expenditure equity. The study found that between FY 2000 and FY 2011 the coefficient of variation ranged from 0.11 to 0.14; the McLoone Index ranged from 0.89 to 0.94 during that time frame; and the Verstegen Index ranged from 1.09 to 1.03. An analysis of the correlation between education spending and local property wealth showed a negative correlation prior to implementation of Act 68 and a positive correlation following its passage. The correlation has increased somewhat in recent years. The 2014 session of the legislature reviewed options to the current finance system. These focused mostly on limiting local spending choices, something that is likely to improve the state’s already strong equity performance (Picus, 2014).

These and the national studies cited above focus on within-state fiscal disparities. Corcoran and Evans (this volume) and Odden and Busch (1998) show, however, that after adjusting for cost differences, the majority of fiscal differences are caused by interstate rather than intrastate disparities. Thus even if all within-state disparities are eliminated, more than two-thirds of the disparities will remain. These studies, however, leave unanswered the question of whether additional money has improved student learning and performance. Understanding this link has been difficult. Below, we briefly discuss efforts to make the use of educational resources more efficient and, following that, suggest how the growing adequacy movement may provide a path for improving student outcomes across all types of districts and among all students.

### Adequacy

In recent years, many states have conducted adequacy studies, and in many instances these studies have led to efforts to raise spending in those states. The Great Recession of 2008 slowed progress in most states, but by 2013, with improvements in the economy and state revenues, there have been renewed efforts to increase spending for schools. Wyoming, Arkansas, and Washington are instructive of possible future trends in school finance adequacy, while California offers an example of a state that has begun to increase education spending focused on students identified as needing the most assistance.

### Wyoming

Wyoming has perhaps more experience than any other state in designing and implementing school funding systems based on adequacy studies. Beginning in 1996 with a court ruling (Campbell County School District v. State) that required the state to define a proper education and fund the cost of that education, Wyoming has worked to estimate the costs of an adequate education and distribute funds to districts in a manner that will allow children to succeed.

Through two iterations five years apart, Wyoming policymakers used a professional judgment method to ascertain the level of resources needed to meet the standard established by the court (see Guthrie & Rothstein, 1999; Guthrie et al. 1997). In 2005, and again in 2010, the state used an evidence-based model to reassess the adequacy of its school funding system (Odden et al. 2005; Picus & Odden, 2010). Following the implementation of the 2005 study’s recommendations, Wyoming school funding was among the highest of the 50 states. While the state’s assessment system showed modest gains in student performance, results from the National Assessment of Educational Progress (NAEP), while improving, paint a less optimistic picture.

In response to growing school funding appropriations, the Wyoming Legislature undertook a two-year study of school district resource use. The state asked its consultants to assess the instructional vision or goals of schools in the state and to compare how each school and school district use the resources allocated to them through the evidence-based funding model in relation to the way those funds were estimated. The findings from this work both confirmed the relatively weak gains in student performance (even in a study focused on
improving schools) and showed that most schools and districts continued to utilize resources in the way they were previously allocated in the district, despite low performance gains and an evidence-based theory of action.

Arkansas

In response to court rulings in Arkansas, that state’s legislature undertook an extensive evidence-based adequacy study in 2003 and updated that analysis in 2006. Following completion of the 2003 study (Odden, Picus, & Fermanich, 2003), a special session of the Arkansas Legislature took the school-based funding estimates developed by Odden et al. and converted that figure to a single per-pupil figure for a school district. This was then appropriated and distributed to each school district in the state.

In 2006, Odden, Picus, & Goetz (2006) recalibrated the adequacy study and as part of that work visited a total of 107 schools across the state to ascertain how schools and school districts were using all of their revenues in relation to the adequacy model design. The researchers found that, as in Wyoming, for the most part, schools were not implementing policies that would allocate resources as designed through the adequacy model, suggesting that a more focused set of expectations on the part of the legislature for the use of all school funds (new and existing resources) might be appropriate.

Washington

Research in Washington supports a recommendation for a more focused set of expectations around the use of school funds. Fermanich et al. (2006) conducted a successful schools analysis in the state of Washington as part of a comprehensive adequacy study done for Washington Learns, a legislatively established 18-month comprehensive study of preschool through university education in Washington. Through that analysis, they identified nine successful school districts and 31 schools in those districts that showed dramatic improvement in student performance in recent years. Site visits to the districts and schools sought to understand the instructional goals and resource use decisions of each school. While local decisions varied to some extent, they were all highly correlated with the recommendations of the evidence-based adequacy study that was also conducted for Washington Learns (Odden et al., 2006).

It took Washington several years to revise its funding formulas, but a few years after the adequacy study it redefined “basic education” to reflect the more robust staffing ratios and funding levels recommended in that 2006 study. However, it did not fund those ratios, thus opening the state to a legal challenge because its courts have ruled it the primary duty of the state to fully fund a “basic education” program. Indeed, in 2012 the Washington Supreme Court found the system unconstitutional because it was not funded and gave the state until 2018 to show substantial progress towards full funding, which at that point required more than $4 billion in additional dollars.

California

School funding suffered in California more than probably any state in the country during the Great Recession. Not only did inflation-adjusted spending decrease, but also nominal dollars available for schools actually declined for three years in a row. In 2012, California voters approved Proposition 30 implementing increases in the state income and sales taxes for several years. At the same time, the legislature enacted the Local Control Funding Formula (LCFF) that dramatically changes how resources are allocated to school districts. Under the LCFF, districts receive funding for each student at a base rate. For each at-risk student (English language learners, students qualifying for free or reduced-price lunches, and students in foster homes), districts receive an additional per-pupil allocation equal to 20 percent of the base allocation. In addition, when the proportion of at-risk students exceeds 55 percent of district enrollment, concentration grants equal to 50 percent of the base allocation are provided for the proportion of students exceeding 55 percent. This influx of new resources, with a much stronger focus on at-risk students represents a dramatic change in how California funds its schools.
Because the LCFF program has an eight year phase-in, it is impossible to predict how districts will use these funds at the time this chapter was written. As part of this new funding program, however, the state requires each district to develop a Local Control and Accountability Plan that describes how it plans to use the new funds to serve at-risk students and link its spending priorities to state goals, including deciding how teachers are assigned to schools and establishing goals for growth in student achievement.

In summary, the experience of states faced with establishing school funding systems based on adequacy has been one of finding the additional funding called for by the adequacy studies that were conducted and then hoping that schools and school districts would use those funds in ways that improved student performance. Early models provided considerable local leeway to schools in how they used their educational resources, although recent findings in states with the most experience in school finance funding formulas based on adequacy studies suggest that a more categorical or directive approach to how money is used might receive more consideration from state policymaking bodies, particularly state legislatures, in the future.

Conclusions

This chapter has focused on the use of intergovernmental aid formulas in school finance. We traced the evolution of school funding formulas in the second half of the 20th century and the early years of the 21st century and showed how they have evolved from basic foundation programs to improve school funding equity to programs designed to meet adequacy goals.

In many ways, the evolution of school funding formulas has been circular. That is, if states can design funding formulas that provide each student with the resources they need to meet state performance expectations, the goals of horizontal and vertical equity will be met as well. Moreover, under an adequacy-based funding system, it seems possible to achieve the goals of fiscal neutrality by using combinations of state and local revenues at uniform tax rates to finance adequate funding levels. Adequacy raises the question of whether all the goals established for school finance formulas—equity, efficiency, productivity, and adequacy—can be met through a singular formula. And if formulas can be constructed to meet all of those goals, are they affordable by most states? If funding turns out to be beyond the fiscal capacity of many states, what is the solution? Given large spending disparities across states, is there a federal role in ensuring adequate educational resources are available for all children? And what are the implications of No Child Left Behind’s achievement expectations for children?

As the courts, state policymakers, and education officials strive to improve student learning, developing a better understanding of how educational resources are translated into student learning is critical. We think that developing this understanding will require extensive and hard work in multiple schools across multiple districts and states to fully understand how resources matter over time. Operationalizing that understanding into school funding formulas will remain a challenge for the immediate future.

Appendix

Operation of State aid Formulas

This appendix provides more detail on the operation of the foundation program and the guaranteed tax base formula.

Foundation Program

A foundation program can be represented as follows:

\[ S_{A_i} = E^a - t_{\min} \cdot (TB_i) \]  

(1)
where

\[ SA_i \] is the state aid per pupil in the \( i \)th district

\( E^* \) is the state-determined foundation level of education spending

\( t_{\text{min}} \) is the state-determined minimum tax rate that each district is required to levy

\( TB_i \) is the property tax base or wealth per pupil in the \( i \)th district

For example, if the state foundation guarantee (\( E^* \)) were $2,000 per pupil and the minimum tax rate were 20 mills or 2 percent of property value, Table 17.1 summarizes state aid per pupil to each district (\( SA_i \)):

**Table 17.1 Foundation Program Computation Examples**

<table>
<thead>
<tr>
<th>District</th>
<th>( TB_i ) ($)</th>
<th>Local revenue per pupil ($)</th>
<th>( SA_i ) ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50,000</td>
<td>1,000</td>
<td>1,000</td>
</tr>
<tr>
<td>B</td>
<td>100,000</td>
<td>2,000</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>500,000</td>
<td>10,000</td>
<td>(8,000)</td>
</tr>
</tbody>
</table>

In this example, District B is what is frequently called the zero aid district; that is, it raises exactly the amount of revenue guaranteed by the foundation program and thus receives no state aid. This is an important concept as all districts with property wealth per pupil below the wealth of the zero aid district will receive some state aid, while all those with property wealth above the wealth of the zero aid district will not receive state aid through the foundation program.

A second issue identified in this table is how districts with wealth above the zero aid district should be handled. In the example above, District C raises $8,000 more than the foundation guarantee. The policy question facing the state is whether to recapture some or all of that revenue to distribute to other districts, or to let District C keep all of the additional revenue it collects at the state mandated tax rate.

Finally, state policy must determine whether or not school districts can exceed the minimum tax rate to increase spending, and if so whether or not those tax rate decisions will be equalized.

**Guaranteed Tax Base**

A guaranteed tax base is designed to ensure that any district regardless of wealth will have as much revenue per pupil as the district with the guaranteed tax base, at any given tax rate. The formula for a GTB can be represented as:

\[
SA_i = t_i (GTB - TB_i)
\]  \(2\)

Thus the total revenue for any school district is the sum of locally generated revenue plus its state aid as follows:

\[
t_i (TB_i) + t_i (GTB - TB_i) = t_i (GTB)
\]  \(3\)

Unlike a foundation program where the amount of aid a district receives is fixed based on the state’s policy decisions (foundation level and minimum tax rate), under a GTB a district can spend any level it chooses depending on the tax rate it establishes.

The GTB is the equivalent of the zero aid district identified in the foundation program above. That is, the district with a wealth equal to the GTB receives no state aid, and for districts with property wealth above the GTB, the same questions about recapture that were identified in the foundation program description must be considered.

For those districts with wealth below the GTB, this program is effectively a matching program. If one represents total spending in the \( i \)th district as \( E \), and assuming a district spends all its revenue, then:
\[ E_i = t_i(GTB) \]  

and  

\[ t_i = \frac{E_i}{GB} \]

Substituting (5) into (3), we find:

\[ SA_i = (1 - TB_i/GB)E_i \]

On the surface, this would appear to give all districts the same tax base (at least those with wealth up to and including the GTB). As Feldstein (1975) showed, that will not lead to the same level of spending in all districts due to a price effect (i.e., the variation in the “cost” to local taxpayers of providing one more dollar of spending to education will vary depending on property wealth of the district). Moreover, as described above, the result of this price effect is not completely predictable. While we would expect low-wealth districts to increase their tax rates substantially because the state share of each dollar raised is high, in a number of states (Wisconsin and Missouri, for example) we have found that it is the high-wealth districts that raise taxes to remain high spending while the low-wealth districts elect to reduce taxes and remain relatively low spending.

Notes

1. An exception occurred in 2009–10 and 2010–11 when federal stimulus funding, combined with reduced state and local revenues, increased the federal share to nearly 13 percent.

2. Although some districts are dependent on their local governments for revenue, much of the local government revenue also comes from property taxes (McGuire, Papke, & Reschovsky, this volume).

3. See Odden and Picus (2014) for more detailed descriptions of how these formulas work.

4. The equations presented here are taken directly from Odden and Picus, 2014. See the Appendix for a more detailed description of these formulas.

5. The categorical program for free and reduced price lunch was actually more complex. The $480 per free and reduced price eligible pupil was for all schools with less than 70 percent of their enrollment qualifying for free and reduced price lunches. For schools with between 70 and 90 percent eligible students, the categorical grant was $960 per eligible student, and for schools with over 90 percent free and reduced price lunch students, the categorical grant was $1,440 per eligible student.

6. Special education programs for children with mild and moderate disabilities were resourced through local school sites, while services for children with severe disabilities were funded 100 percent by the state. In the actual funding model, all resources were converted into a per-pupil foundation level, and in the case of special education, three teacher positions were added to the number of teachers required for a school with 500 students and funded at the statewide average total compensation. These three teacher positions are included in the $5,400 per student foundation level.

References


McInnis v. Shapiro, 293 F. Supp. 327 (N.D. Ill. 1968), aff’d.


Picus, L. O. (2014). *Funding Vermont’s schools for the future: A policy brief prepared for the Shumlin*


Introduction

All countries face issues of educational equity. Depending on the country, the policy debate may focus on how to increase access to primary or secondary schools; how to reduce persistent achievement gaps between students of different genders, ethnic backgrounds, or income levels; or how to reduce educational resource disparities between rural and urban areas of the country. Likewise, educational priorities in countries where a high proportion of available jobs require high levels of skills will differ from those with less demanding job markets. Though all such issues are of interest to policymakers and researchers, we focus in this chapter on the major areas of educational policy and research at the international level aimed at promoting equity in education, particularly in developing countries.

Specifically, we begin with the international Education for All (EFA) movement whose principal goal is to promote access to quality education for all students around the world. We then turn to two specific topics that have moved to the center of the global agenda: efforts to promote gender equity by expanding educational opportunities for girls and efforts to promote fiscal equity by eliminating school fees. Both goals are not only worthy in themselves but are also keys to promoting other objectives such as economic development that are important to the international community. A leitmotif that runs through all three sections is the trade-off that policymakers frequently confront between the competing values of greater access and enhanced quality.

Education for All

A global consensus has emerged around a number of goals aimed at ensuring that each country provides all of its children with a quality basic education. The most notable expression of this consensus has been the Education for All movement that grew out of the World Conference on Education for All in Jomtien, Thailand, in 1990.

Genesis of EFA

James Grant, the administrator of the United Nations Children’s Fund (UNICEF), proposed holding a conference that would produce a worldwide commitment to meeting the learning needs of children, youth, and adults. He enlisted the support of the United Nations Educational Scientific and Cultural Organization (UNESCO), the Educational and United Nations Development Program (UNDP) and the World Bank to organize the project. Delegates from 155 member states of the United Nations and more than 150 organizations attended the conference and adopted a World Declaration on Education for All (henceforth the Jomtien Declaration) in which they pledged “to act cooperatively through our own spheres of responsibility, taking all necessary steps to achieve the goals of education for all.” They also agreed on a Framework for Action that laid out strategies by which national governments, international organizations, bilateral aid agencies, nongovernmental organizations and others could pursue these objectives (UNESCO, 1990).

The Jomtien Declaration was reaffirmed and updated a decade later, in 2000, by participants from 164 countries at the World Education Forum in Dakar, Senegal. They adopted a Dakar Framework for Action
(henceforth the Dakar Framework) that laid out six specific goals for EFA, three of which set target dates (see box) (UNESCO, 2000a). The coordinating agency for the global effort to implement Education for All is UNESCO.


1. Expanding and improving comprehensive early childhood care and education, especially for the most vulnerable and disadvantaged children.
2. Ensuring that by 2015 all children, particularly girls, children in difficult circumstances, and those belonging to ethnic minorities, have access to and complete free and compulsory primary education of good quality.
3. Ensuring that the learning needs of all young people and adults are met through equitable access to appropriate learning and life-skills programs.
4. Achieving a 50 percent improvement in levels of adult literacy by 2015, especially for women, and equitable access to basic and continuing education for all adults.
5. Eliminating gender disparities in primary and secondary education by 2005, and achieving gender equality in education by 2015, with a focus on ensuring girls’ full and equal access to and achievement in basic education of good quality.
6. Improving all aspects of the quality of education and ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills.

Much of the impetus behind the EFA movement grew out of recognition that the prevailing model of piecemeal, donor-driven projects to improve education in developing countries had been unsuccessful (Schubert & Prouty-Harris, 2003). Such projects, typically, were too narrow in scope, did little to build local capacity to sustain educational improvements, and were designed with the priorities of donors rather than recipients in mind.

**The Conceptual Basis of Education Equity in EFA**

The conceptual starting point for the EFA movement rests on three pillars: human rights, economic development, and poverty alleviation. The Universal Declaration of Human Rights, adopted by the United Nations in 1948, asserted that “everyone has a right to education,” and this principle has been reiterated in numerous subsequent international documents, including the 1979 Convention on the Elimination of All Forms of Discrimination Against Women and the 1989 Convention on the Rights of the Child. The payoff in terms of the economic development that ensues from investing in basic education in developing countries has been widely documented especially for girls (Psacharopoulos & Patrinos, 2002). Finally, investment in basic education is widely viewed as an essential tool in efforts to alleviate poverty.

As part of its global antipoverty strategy, the United Nations General Assembly in 2001 adopted a set of five Millennium Development Goals. Two of them—achieving universal primary education by 2015 and eliminating gender disparities in primary and secondary education by the same year—were education-specific and overlapped with the Dakar Goals (UNESCO, 2005, p. 28). UNESCO literature cites additional reasons for the international community’s stake in educational equity, including global political and economic stability (UNESCO, 1997).

**Defining EFA: Access Plus Quality**

From the outset, proponents of EFA rejected a narrow conception of “basic education” as measured by a certain number of years of schooling, and opted instead for an inclusive definition. The Jomtien Declaration
characterizes basic education as the knowledge and skills “required by human beings to be able to survive, to
develop their full capacities, to live and work in dignity, to participate fully in development, to improve the
quality of their lives, to make informed decisions, and to continue learning.” It adds that the scope of basic
learning needs and how they should be met varies across countries and cultures and changes over time
(UNESCO, 1990). The expanded list of six goals adopted at Dakar was described as embracing “a holistic
conception of educational development” (UNESCO, 2005, p. 28).

Jomtien inspired at least three major projects aimed at improving the quality of education in developing
countries, including the Monitoring Learning Achievement project, initiated by UNESCO and UNICEF, to track
learning achievement around the world at the fourth-grade level (UNESCO, 2000a). At the behest of several
Southern African countries, the Southern and Eastern Africa Consortium for Monitoring Educational Quality
(SACMEQ) was created in 1995 to monitor progress towards EFA quality goals and to develop local capacity to
measure student achievement (SACMEQ, 1993).

Despite such programs, most EFA efforts in the early 1990s focused on the relatively narrow goal of
expanding access to formal primary education, with progress measured in quantitative terms by trends in
school attendance and rates of persistence. Little and Miller (2000) argue that the World Bank, UNICEF, and
others saw this focus on access as a matter of technical and political convenience, and some critics viewed this
initial emphasis on access as a retreat from the richness of the vision of Jomtien. Just as important as getting
students into schools, they argued, was the quality of instruction that they received once they were there
(Delors et al., 1996). Such concerns were voiced at the 1996 Amman Forum on EFA, which was called to review
progress since Jomtien, and by the mid-1990s the EFA movement was actively seeking a balance of access and
quality (UNESCO, 1996). Some observers, however, believed the focus on access was appropriate. Carnoy (2004,
p. 19) argues, for example, that "some countries have such low levels of education ... that simply increasing the
average number of years of education attained may still be the most efficient strategy to follow." In 2011 the
World Bank published a new set of strategic goals for education through 2020 that put a new twist on the
meaning of quality. The document, entitled "Learning for All," argued that "the overarching goal is not just
schooling, but learning." While getting millions more children into school has been a great achievement," it
said, the overriding focus should now be on what individuals actually learn, both in and out of school (World

The current renewed emphasis on quality is attributable in part to the relative success of efforts to expand
primary enrollments. According to the 2012 Global Monitoring Report, for example, by 2010 net primary
enrollment rates (see definition below) had increased in nearly every country whose enrollment rates at the
beginning of the 1990s had been below 70 percent. The number of countries with primary net enrollment ratios
of over 97 percent increased from 37 to 55 out of 124 countries between 1999 and 2010 (UNESCO, 2012, p. 58).
Such successes, however, came with a price in the form of trade-offs between access and quality. In some
countries and regions, particularly sub-Saharan Africa, the rise in enrollments strained resources and resulted
in substantially higher pupil-teacher ratios (UNESCO, 2004). In 2000, the expanded commentary on the Dakar
Framework cited “a lack of attention to the quality of learning” as one of several reasons for the failure to
achieve the vision of EFA established at Jomtien (UNESCO, 2000a). Nonetheless, it is noteworthy that the two
Millennium Development Goals relating to education—universal education and eliminating gender disparities
—stress access rather than quality.

The 2012 EFA Global Monitoring Report also notes that since 2000 most attention has focused on the three
Dakar goals that are directly related to formal elementary education systems (universal primary education,
gender parity, and quality of education) while the other three (early childhood care and education, the learning
needs of young people and adults, and adult literacy) have received relatively little attention. It described adult
literacy in particular as "the goal that is probably the most neglected in the ERA agenda." (UNESCO, 2012, p.
167). The report points out that achieving the latter three goals typically requires the creation of a set of new
programs that do "not fall neatly under the mandate of the Ministry of Education but rather is spread among
several ministries" (UNESCO, 2012, p. 28). It also notes that governments tend to think of investments in young
children as producing the greatest educational and political returns.

Measuring Progress Toward EFA
One reason that access received particular emphasis in the early 1990s was that it was easier to measure than quality. A major factor in the widespread neglect of the goal of adult literacy is that it has been framed as a "vague aspiration, with no quantifiable benchmarks against which progress could be measured" (UNESCO, 2012, p. 171). Even today international organizations are struggling to develop quantitative measures of educational quality.

**Measures of Access.** The fundamental goal of universal primary education (UPE) is typically measured by enrollment as captured by two ratios. The gross enrollment ratio (GER) compares the number of children enrolled at the primary or secondary school level, regardless of their age, to the number of children "officially corresponding to that level" (UNESCO, 2005, p. 420). This ratio, expressed as a percent, can exceed 100 if there is early or delayed enrollment or grade repetition in an educational system. The net enrollment ratio (NER) reflects the percentage of age-appropriate children in the population who are enrolled in either primary or secondary school. A NER of 100 percent means that all eligible children are enrolled in school. UNESCO has also measured access through the survival rate to fifth grade. A combination of a high NER and high survival rate to fifth grade implies a high rate of primary school completion (Hewett & Lloyd, 2005).

Problems associated with these measures include the fact that NERs are not entirely comparable across countries due to variation in length of their primary school cycles (five to seven years) as defined by UNESCO. Countries with longer cycles may have more difficulty reaching high NERs. Questions have also been raised about the quality of the underlying data because of poor information systems in some countries and financial incentives for schools to inflate enrollment estimates (Hewett & Lloyd, 2005).

**Measures of Quality.** Compared to access, educational quality is far harder to define and measure. The prevailing understanding of quality education reflected in UNESCO documents derives from the 1989 Convention on the Rights of the Child (CRC), Article 29 of which defines five proper aims of education. In brief, these are "the development of the child’s personality, talents, and mental and physical abilities," "the development of respect for human rights and fundamental freedoms," "the development of respect for the child’s parents and her own and others’ cultures, "the preparation of the child for responsible life in a free society," and "the development of respect for the natural environment." None of these is easy to quantify.

The Framework for Action that accompanied the Jomtien Declaration tried to address the quality issue by specifying a target of "improvement in learning achievement such that an agreed percentage of an appropriate age cohort (e.g., 80 percent of 14-year-olds) attains or surpasses a defined level of necessary learning achievement" (UNESCO, 1990, p. 18). That approach, however, turned out to be unworkable because very few countries had set—much less met—achievement targets (UNESCO, 2000a). Thus, it was not until Dakar that UNESCO developed a working definition of quality by which countries’ progress could be measured. The expanded commentary on Goal 6 of the Dakar Framework relating to quality education lists eight requirements of successful education programs, including well-trained teachers, active learning techniques, participatory governance and management, and a "relevant curriculum that can be taught and learned in a local language." The Framework also laid out a specific overall target: "Improving every aspect of the quality of education and ensuring their excellence so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy, and essential life skills" (p. 17).

Despite the above reference to learning outcomes, it is significant that most of the measures used to define quality focus on inputs such as qualified teachers rather than on learning outcomes. Comparisons are complicated by the fact that many developing countries do not participate in major international tests such as Trends in International Mathematics and Sciences Study (TIMSS) and the Program for International Students Assessment (PISA) because of insufficient technical capacity or other reasons. Efforts to refine the definitions and means of measuring quality education at the basic level are continuing. The 2005 Global Monitoring Report, subtitled "The Quality Imperative," uses widely available input measures of educational quality and, to the extent possible, internationally comparable assessments of cognitive ability to assess the state of school quality in developing countries (UNESCO, 2004).

**Progress in Access and Quality**

In adopting the Dakar Framework for Action, the World Education Forum agreed to report progress toward the six goals. To facilitate such reporting, UNESCO in the 2003/4 Global Monitoring Report introduced the
Education for All Development Index (EDI), a composite measure of a country's progress toward providing a basic, quality education to all its citizens. The EDI focused on four of the six goals: universal primary education, adult literacy, gender parity, and the quality of education. It excluded Goal 1 (early childhood care and education) on the grounds that reliable and comparable data are not available for most countries, and Goal 3 (learning needs of youth and adults) because this objective "has not yet been sufficiently defined for quantitative measurement" (UNESCO, 2005, p. 252).

Proxies were chosen for each of the four goals. Universal primary education (Goal 2) is measured by the primary net enrollment rate and adult literacy and (Goal 4) by the literacy rate among individuals 15 years and older. The proxy for gender parity (Goal 5) is a composite Gender Parity Index based on the arithmetical mean of primary and secondary gross enrollment ratios for boys and girls and the adult literacy rates for women and men. In the absence of data on learning outcomes comparable across countries, the proxy for quality of education (Goal 6) is the proportion of students surviving to primary grade five. For purposes of compiling an overall EDI score, each of the four goals was given equal weight. The index runs from 0 to 1, with countries having an index score of 0.95 judged to have achieved EFA or being close to doing so.

The 2012 Global Monitoring Report includes EDI scores in 2010 for the 120 countries for which data were available on all four of the component measures (UNESCO, 2012, Table EDI.I.1, pp. 308–309). Of these, 58 countries were found to have reached the 0.95 threshold signaling that they have achieved EFA. A majority (60 percent) of these countries are in either Central and Eastern Europe or North America and Western Europe. Forty-two other countries, mostly in the Arab States and Latin American and the Caribbean, had EDI values between 0.80 and 0.94, while 20 countries, 13 of them in sub-Saharan Africa, had values below 0.80, which meant that "achieving EFA would require intervention throughout the school system." The 2012 GMR also includes an "extended" EDI that includes a fifth component reflecting an Early Childhood Care and Education (ECCE) index for 52 countries for which this index can be calculated. Some countries that have put more emphasis on early childhood, especially in Central and Eastern Europe and in Latin America and the Caribbean, saw their standing on the EDI rise considerably.

With regard to the six specific Dakar goals, the 2012 GMR reports that early childhood enrollment (Goal 1) is improving, albeit from a low base in some regions. Nevertheless, despite a 46 percent increase in the number of children enrolled in preschool between 1999 and 2010, "less than half the world’s children receive preprimary education." Based on current trends—and despite the fact that the number of out-of-school children of primary school age fell from 108 million in 1999 to 61 million in 2010—the target of universal primary education by 2015 (Goal 2) will be missed. The number of children enrolled in secondary school has increased by 25 percent since 1999, but 12 years after Dakar, the international community is only now coming close to agreement on a coherent set of indicators to measure skills development (Goal 3). With regard to adult literacy, the GMR reports that "most countries will miss Goal 4 by a wide margin." There were still 775 million adults who could not read or write in 2010, about two-thirds of whom are women. By contrast, Goal 6 relating to gender parity and equality "has been one of the successes of the EFA movement since 2000, but more needs to be done" (p. 106).

Gender Equity

Although multiple forms of internal inequities characterize the education systems of developing countries (Wils, Carrol, & Barrow, 2005), only one particular form of inequality—that between genders—is singled out in both the Dakar Framework and the Millennium Development Goals. The Dakar Framework established two specific targets related to gender equity: (1) the elimination of all gender disparities in enrollment in primary and secondary education by 2005, and (2) gender equality in education by 2015. The 2005 Global Monitoring Report, which focuses on progress toward these goals, discusses these targets from two perspectives: fundamental human rights and economic development.

The educational rights perspective draws heavily on a background paper by Duncan Wilson, later published in the journal Prospect (Wilson, 2004). Wilson grounds his rights argument in the widespread ratification of the 1989 Convention on the Rights of the Child and the 1979 Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW), both of which assert the right to gender equality in education, and
he makes a distinction between parity and equality. Gender parity, in Wilson’s view, refers to equal access to education and is easily captured in quantitative measures such as enrollment rates. By contrast, educational equality encompasses rights “in” and “through” education, such as equal treatment within schools and equal opportunities for developing one’s talents, and is qualitative in nature. This distinction between parity and equity has clear parallels to the more general EFA discussion of the differences between access and quality.

The economic arguments for the importance of gender equality in education generally come in two varieties: those based on the private returns to education for girls and women and those based on positive externalities associated with expanding education for them. Studies of the private returns to education by King and Hill (1993), Psacharopoulos and Patrinos (2002), and Schultz (1995) have found that even though women on average earn less than men of the same age and education level, the return to a year’s education for a woman is equal to, or slightly greater than, that for a man. A more persuasive argument for equal access for girls rests on evidence that the private returns on investments in education at the primary level are higher than those for higher levels of education (Psacharopoulos & Patrinos, 2002; Schultz, 2001). It follows that in countries where girls have little access to primary and basic secondary education, expanding their access to education would be a more efficient use of resources than expanding education without changing the gender gap. Such returns to education are “private” in the sense that they are captured in wage rates and accrue directly to the participants. By educating girls, a country makes more productive use of its resources and thereby increases average income levels.

Studies also document significant positive externalities associated with investment in expanding education for women. Schultz (2002) cites studies showing that increases in a mother’s schooling are associated with greater improvements in outcomes for a family’s children than are increases in the father’s schooling. These outcomes include birth weight, child survival, nutrition, age at which schooling begins, and years of schooling by adulthood. Schultz also references the consistent finding that fertility is inversely related to women’s schooling.

Whereas Schultz focuses on microeconomic evidence of the benefits of expanding schooling for women, others have tried to model the effect of gender equality in education on macroeconomic growth. Klasen (2002) found that 0.4 to 0.9 percentage points of the differences in annual per capita gross domestic product (GDP) growth rates between East Asia and sub-Saharan Africa, South Asia, and the Middle East over the period from 1960 to 1992 are attributable to differences in the gender gaps in education between the regions. This effect is estimated controlling for the level and growth of human capital.

Based on the literature, Abu-Ghaida and Klasen (2004) estimated the cost in foregone economic growth in 45 countries unlikely to meet the Millennium Development Goal (MDG) of eliminating the gender disparity in primary and secondary education by 2015. They conclude that countries significantly off the track of meeting this goal could lose 0.1 to 0.3 percentage points in annual economic growth for the decade from 1995 to 2005 and an additional 0.4 percentage points annually from 2005 to 2015. Additionally, they estimate that these countries will record 0.1 to 0.6 more children per woman by 2015 and 32 more child deaths per 1000 children than they would have had they met the MDG of gender equity. All such empirical studies are subject to criticism because of the weaknesses of the quantitative measures used to test the theoretical models on which the empirical work is based, and the difficulties of controlling adequately for confounding factors.

Obstacles to Gender Equity

Impediments to equality in education for women are numerous. They range from early marriage and pregnancy to adverse treatment when they do enroll in school in forms such as biased textbooks, a dearth of female teachers, lack of sanitary facilities, and sexual harassment. In addition, due to their lower social status, girls in many countries are more vulnerable than boys to factors that reduce overall access to education. Faced with school fees and other cost pressures, for example, families are less likely to enroll a female child than they are a male child. Girls frequently bear the burden of caring for siblings or for parents and others with HIV/AIDS. Social norms and cultural expectations limit girls’ life trajectories in the workplace.

Most of the evidence of early marriage and pregnancy as barriers to educational access for women is qualitative and country or region specific. CEDAW reports from Egypt and Morocco mention early marriage as a cause of female drop-out in rural areas. They also provide the fullest documentation of the role of early
pregnancy in preventing women from receiving a quality basic education (Wilson, 2004). The nature of the relationship between marriage or early pregnancy and education may not be causal, of course; rates of early marriage are likely to be higher, and girls’ participation in school is likely to be lower, in countries and regions with social norms that promote women’s domesticity. But it is also the case that women in some countries are withdrawn from school specifically to marry (UNICEF, 2001).

Human rights reports have documented the prevalence of in-school abuse of girls in South African schools, and case studies of violence in schools in Jordan, Ethiopia, Malaysia, Israel, Slovakia, and several South American countries document that girls are disproportionately victimized by in-school violence in most of these countries (Ohsako, 1997). Wilson (2004) cites the reports of Thailand and the Philippines from 1997 and 1996, respectively, as evidence of bias in curriculum.

**Progress Toward Gender Equity**

The 2012 Global Monitoring Report describes the narrowing of the gender gap in primary enrollment as “one of the biggest EFA successes since 2000,” but it adds that “some countries are still in danger of not achieving gender parity in primary and secondary education by 2015” (UNESCO, 2014, p. 308). Sixty-eight countries have not yet achieved gender parity in primary education, but the incidence of severe gender disparity has become less common. Of 167 countries with data in both 1999 and 2010, the number of countries with fewer than nine girls in primary school for every 10 boys dropped from 33 to 17. The number of countries where girls face extreme disadvantage—defined as gender parity below 0.70—has fallen consistently from 16 in 1990 to 11 in 2000 to only one in 2010 (Afghanistan). The key reason for girls not being in school is not starting school in the first place (UNESCO, 2012, pp. 106–107).

Gender disparities are also narrowing at the secondary level. Among countries with data for both 1999 and 2010, the number with fewer than nine girls enrolled for every 10 boys has declined from 28 to 22, a majority of them in sub-Saharan Africa. The GMR emphasizes that gender equality—defined as “providing a safe, secure and supportive learning environment for all” as well as attention to learning outcomes—often remains a challenge even when gender parity is realized. In general, girls perform better than boys in reading at both the primary and secondary school levels, and these gaps are widening. Boys have an advantage in mathematics in most countries, but this gap may be narrowing. Several new initiatives have highlighted the issue of gender parity and equality. In May 2011 UNESCO launched both a Global Partnership for Girls’ and Women’s Education and a High Level Panel on Girls’ and Women’s Education for Empowerment and Gender Equality.

Whereas the typical challenge for gender equity is to increase access to schooling for girls, the 2006 Global Monitoring Report also noted that in a growing number of countries the enrollment rates for girls are higher than for boys, especially at the secondary and tertiary levels. This pattern is typically found either in developed countries or in developing ones that are close to achieving universal primary education. For this reason, some developed countries, such as Denmark, Finland, New Zealand, and the United Kingdom, are at risk of not achieving gender parity in secondary education by 2015.

**Funding Equity: School Fees**

One of the thorniest issues related to funding equity in developing countries is the balancing of public and private funding in school finance. In contrast to practice in the United States and other developed countries where public schools are financed almost entirely with public revenue, families in developing countries are often asked to assume responsibility for a significant share of school funding through enrollment fees paid directly to schools or other schooling-related expenditures, such as textbooks and compulsory school uniforms. According to the 2010 Global Monitoring Report, at least 55 countries have no legal guarantee to free primary education, with almost half of them being in sub-Saharan Africa (UNESCO, 2010, p. 332). However, even when they are not officially permitted, many schools still charge fees in some form.

Despite their prevalence, the international community is increasingly opposed to school fees, especially at the level of primary education. Critics argue that such fees keep some children from attending school, an
unacceptable outcome for a service deemed by the international community to be a basic human right. Further, because a major purpose of primary education is to reduce poverty, a related concern is that, since school fees are likely to impose the greatest burdens on poor households, they are thus inherently unfair. Notably, the Dakar goals explicitly call for free and compulsory education of good quality.

Given the relatively large role that user fees currently play in the financing of education in many developing countries, their elimination raises a number of complex policy challenges that revolve around the access vs. quality dilemma. When fees are removed, alternative sources of revenue must be found not only to replace the revenue that would have been generated by school fees but also to cover the costs of sustaining quality for a larger group of enrolled students.

The Prevalence of User Fees for Primary Education

The most comprehensive information on school fees comes from a 2004 World Bank survey covering 79 of its 145 member countries (Kattan & Burnett, 2004). This survey indicates that, with the exception of Algeria and Uruguay, all of them impose some form of fee on parents of public school students. In about a third of the countries, however, the fees are not sanctioned by the government and are, in fact, counter to national policy. Reported user fees include payments for tuition in the form of school registration and annual fees (30 countries), textbook charges (37), compulsory uniforms (39), PTA/community contributions (56), and activity fees (34) (Kattan & Burnett, 2004, p. 10). For the purposes of this survey, fees do not include payments for school meals or for private tutors.

The prevalence by type of fee varies across regions. Tuition fees are quite common in the former socialist countries and East Asia, including China, but are almost nonexistent in Latin America. Fees for textbooks are the norm in all areas, but compulsory uniforms are much more prevalent in the former socialist countries than elsewhere. All types of fees are quite common in Africa, including contributions to community schools.

Data from 12 developing countries derived from household surveys in the 1990s indicate that household spending on a broad array of school-related items accounted on average for over 20 percent of total public and private spending on primary education. For example, families often see a need to purchase private tutoring from teachers in order to ensure that their children will receive the attention they need in situations where there are large classes. In Cambodia more than a third of the household expenditure goes to private tutoring, while in India about three-quarters goes toward textbooks and other learning materials.

In some countries, including China and South Africa, school fees vary in magnitude across schools and hence contribute to disparities in school quality. In China, high-quality schools in urban areas, known as key schools, often recruit fee-paying students from other districts. Such fees, which are not allowed under national law, are often described as "education donation and assistance" or "joint construction fees" (World Bank, 2005, p. 3). In South Africa, data from one of the wealthiest of its nine provinces indicate that the fee revenue collected by schools in 2002 serving middle-class students—both white and black—permitted those schools to increase their teaching staffs by more than 30 percent on average above their state-paid allotment of teachers and in some schools to double the number of teachers they could employ (Fiske & Ladd, 2004). Although legally required to provide fee waivers to low-income children, schools that opt for high fees had strong incentives to discourage such children from enrolling.

Although poor schools in South Africa at that time were also permitted to impose fees, their fees tended to be very low and often went uncollected. In recognition of that fact, South Africa changed its policy in 2006 by providing extra public funding to the poor schools in exchange for giving up the power to levy fees. Nonetheless, the persistent use of fees by middle-class schools continues to generate funding disparities across schools.

Effects of User Fees on Enrollment

Fees differ from other types of funding for education in that they are triggered only when a child is enrolled in school. Because parents can avoid paying school-related fees by not sending their child to school, there is widespread concern that the use of fees will reduce school enrollment rates, especially among poor families.
Evidence from countries such as Bulgaria, Russia, Azerbijan, Armenia, Tajikistan and the Kyrgyz Republic, Moldova, Romania, and Hungary provide some support for this concern (Kattan & Burnett, 2004, p. 18).

By far the most compelling evidence that fees reduce enrollments emerges from the experiences of several African countries that abolished school fees as part of their effort to promote universal primary education. In Malawi, the 1994 elimination of school fees generated a 50 percent increase in primary school enrollment, from approximately 1.9 million in 1993–94 to nearly three million the following year (Kadzamira & Rose, 2001). Similarly, when Uganda eliminated tuition fees for primary schools in 1996, school enrollment increased by 70 percent in one year and rose an additional 20 percent by 2001, with the gross enrollment rate at the primary level rising to 123 (Hillman & Jenkner, 2002; World Bank, 2002). Apparently, the elimination of school fees attracted a large number of overage children back into the system. In Kenya, the response to the 2003 elimination of fees was an immediate increase of 1.2 million students, many of whom could not immediately be accommodated.

The Kenyan experience has attracted significant research attention because one surprising outcome of the removal of fees was a large surge in private school enrollments. Some researchers attribute this increase not to reduced per-child resources in the public schools but rather to their changing mix of students. The influx of poor students to the public schools induced affluent students to avoid such schools in favor of private fee-paying schools, thereby increasing socioeconomic segregation within the school system (Bold et al., 2011). Other researchers attribute the surge in private enrollments to the declining quality of the public schools as evidenced by higher pupil-teacher ratios (Nishimura & Yamano, 2013). The most detailed study observes that the private school surge was not limited to nonpoor students (Oketch et al., 2010). This study of the removal of school fees in Nairobi shows that in the slums, but not in other areas, it was the poorest students who ended up in low-cost private schools largely because the government failed to provide sufficient public school spaces for poor children in those areas.

Although some researchers conjecture that the price elasticity of demand for schooling in developing countries is greater for girls than for boys, a recent careful empirical study of school fees in Uganda concluded that the price elasticities were approximately the same: –0.32 for boys and –0.34 for girls, with the difference not statistically significant (Lincove, 2012). If this finding holds true in other contexts, it implies that gender-neutral policies to reduce school fees by themselves are not likely to reduce the gender gap in access.

More generally, researchers agree that simply eliminating school fees is unlikely to generate universal access to primary schools because other factors also keep poor children out of schools, including the costs of transportation, tutoring, and uniforms. In a 2009 study, researchers found that reducing costs to families in Kenya by providing free uniforms reduced student absenteeism by 44 percent (Evans et al., 2009). Other obstacles to schooling include the opportunity costs to families of having children in school rather than working in the home or fields, expectations of low economic returns from a primary education, social norms that reduce benefits to girls attending school, the burdens families face in dealing with HIV/AIDS, and the absence of high-quality schools in rural areas. For very poor families, the greatest obstacle may simply be the absence of resources that would allow them to deal with the opportunity costs inherent in sending their children to school. Consistent with that conclusion is that the Progresa program in Mexico, which gave cash grants to families conditional on their sending their children to school, increased student enrollment among low-income families (Kremer, 2003). Indeed, in some cases policies such as conditional cash transfers may do more to increase school access and reduce disparities than policies that reduce school costs (Lincove, 2009 and 2012).

Even if school fees do not inhibit enrollment, they are undesirable to the extent that they impose financial hardships on poor families. Information gathered by a household survey by income quintile for eight Asian countries document that in many countries not only do school fees place a heavy burden on households in general, but the burdens are highest for families with the lowest income. In China, for example, while all households spend an average of 19 percent on primary education, the proportion for the poorest quintile is 29 percent (Kattan & Burnett, 2004). Thus, user fees for education force poor families to confront difficult financial trade-offs.

Both because user fees tend to reduce enrollment and because they put the heaviest burden on low-income households, the consensus in the international community is that school fees are undesirable and should be eliminated. If basic education is truly a fundamental human right, the argument goes, it is inappropriate to
rely, even if only partially, on a financing mechanism that denies that right to some children. Similarly, if education is intended to play an essential role in reducing poverty, any financing mechanism that disproportionately burdens the poor should be avoided. This consensus against school fees is embodied in Goal 2 of the Dakar Framework, which calls for free and compulsory primary education for all students by 2015.

Why Fees Are So Common

Despite these concerns about school fees and despite the opposition of the international community, many developing countries continue to rely on such fees and even to increase such reliance or to make explicit policy decisions to allow schools to impose fees. They do so primarily because of their inability to generate sufficient public funds—either because of lack of revenue-raising capacity or lack of political will—to provide high-quality basic education for all their students.

A second reason for the prevalence of school fees is that until quite recently some international development organizations, including the World Bank, promoted the use of cost-sharing methods in social sectors, including education, as part of their structural adjustment policies aimed at making developing countries become more fiscally responsible. In addition, the use of cost sharing arrangements—such as having parents pay for textbooks—was rationalized in part on the ground that they would generate greater buy-in from parents and would, by analogy to the private sector, lead to more efficient operation of the schools. We know of no research, however, that documents these alleged beneficial properties of school fees, especially in schools serving low-income families, and in some cases such policies have been clearly counterproductive. In Tanzania, for example, a persistent economic crisis forced the country to negotiate with the International Monetary Fund (IMF) for a loan, conditioned on the country pursuing a number of structural adjustment policies. Among these was the introduction of user fees in the education sector, which forced it to reintroduce school fees which had earlier been eliminated. The predictable result was that gross rates of enrollment at the primary level declined from about 90 percent in the early 1980s to between 66 and 75 percent a decade later (Vavrus, 2005, p. 182). It is of interest that the World Bank made an about-face on the issue of school fees and no longer promotes their use (Kattan & Burnett, 2004).

A third justification for school fees arose most clearly in South Africa at the end of the apartheid era in the mid-1990s. The new post-apartheid government chose not only to permit, but also to encourage, the governing bodies of each school to levy and collect school fees from the families of its students. Not surprisingly, the schools serving middle-class students were able to levy and collect far more fee revenue than those serving poor students. Such disparities were rationalized primarily as a way to keep middle-class families—both black and white—in the public school system and thereby to avoid the situation in many Latin American countries where the middle class has fled the public system (Fiske & Ladd, 2004). A secondary goal was to encourage parental involvement and promote a sense of ownership in the schools, especially among poor black families who had become alienated from the public school system during the apartheid period. Though the policy was relatively successful in maintaining the quality of the middle-class schools and thereby keeping middle-class families of all races in the public schools, the policy also served to perpetuate many of the earlier disparities in quality across schools, with the significant difference being that the new fault line has become family income rather than race.

Although the fee policy was modified in 2006 to apply only to the schools serving children from relatively affluent families, with larger government grants to poor schools, insufficient funding remains a serious problem for many South African schools.

The Quality Challenge

Though the African countries that abolished school fees were remarkably successful in increasing access, this achievement put huge strains on the existing education systems and thereby reduced the quality of education they could provide. This issue was exemplified in Malawi, where rising enrollments in primary schools led to a drastic deterioration in school quality, as measured by stress on school inputs. That country ended up with extremely high ratios of pupils to permanent classrooms (119:1); pupils to desks (38:1); pupils to textbooks
and pupils per teacher (62:1) (Kattan & Burnett, 2004, p. 8). Likewise, in Uganda the number of pupils per teacher rose from 40 before the change to 60 in 1999, while the number of pupils to permanent classrooms rose from 85 to 145 (World Bank, 2005, p. 10).

A comprehensive study of the abolition of school fees in five African countries concluded that “access without quality solves neither the equity program nor the wider objectives of reaching EFA” (International Bank for Reconstruction and Development/World Bank, 2009, p. 14). The case studies highlighted the specific challenges in rural areas where insufficient government funding often led to the creation of community schools supported by parents and external donors and that employed low-paid teachers. The study argued that governments should pay all teachers and provide grants to cover the costs of nonsalary inputs such as textbooks and operating expenditure.

Crucial to the ability of countries to maintain school quality, however, is their ability to generate new funding for education. Such new funding must be sufficient to cover not only the loss of fee revenue from current students but also the significant additional costs associated with the larger number of students. In addition, to the extent that the elimination of fees at the primary level succeeds in increasing primary level enrollments, it also increases the demand for, and, hence the costs of providing, secondary education.

While such additional public funding is clearly a necessary condition for success in meeting the EFA goals of universal free primary education of good quality for all students, it is far from sufficient. As already noted, fees are not the only factor keeping students out of school. Many developing countries currently lack the capacity to train an adequate teaching force and the managerial capacity to run a more extensive system of education.

The two main potential sources of new funds are additional public revenues and additional support from the international donor community. Another possibility is to make schools more productive by means such as reducing retention rates and thus allowing children to take fewer years to progress through school. However praiseworthy that latter strategy might be, it is unlikely to be successful. The strategy itself would require additional resources, and in many low-income areas nonschool considerations such as HIV/AIDS, cultural norms, and the opportunity costs to families of sending children to school rather than working in the fields make it difficult for many children to progress at the expected rate through school. Reducing corruption and misuse of funds, however, could potentially free up some additional resources for education. Uganda’s 1996 Public Expenditure Tracking Survey provides one example of how the provision of information on the flow of funds can reduce the diversion of funds for private or noneducation purposes (UNESCO, 2005, p. 83).

Provided the political will is there, some developing countries appear to have the capacity to generate additional public support for primary education because they currently devote relatively low proportions of their GDP to education. While some countries, including South Africa, Tanzania, Kenya, and Swaziland currently spend 6 percent or more of their GDP on education, most developing countries spend far less.

The harsh fact remains, however, that developing countries are, by definition, very poor and face a number of competing demands on their public budgets for other essential services, such as health and housing. Thus, it may be the case that the only way for many such countries to achieve the EFA goal of free and universal primary education is through substantially more outside support from the international donor community, not only for education, but also more generally to break the cycle of poverty (Avenstrup, Liang, & Nellemann, 2004; Kattan & Burnett, 2004).

**Conclusion**

The Education for All movement has succeeded in ensuring that issues of educational equity remain high on the policy agendas of developed and developing countries alike. Moreover, demonstrable progress is being made in enhancing access to education, thereby moving toward specific goals of universal primary education and gender equity. When it comes to the quality of education, however, the picture is much less clear. Measures of progress are imprecise, and developing countries inevitably face trade-offs between access and quality.

All countries throughout the world struggle with the challenge of providing high-quality education, and none has come up with a silver bullet for doing so. This challenge is particularly daunting in developing countries that are simultaneously attempting to meet the twin goals of universal access to primary education
and increasing the quality of the education they provide. Given their poverty and the size of the challenge, ongoing pressure and enhanced support from the international community—economic, political, and technical—seems essential.

The availability of data at the national level, including achievement data, which are increasingly available through international organizations, will provide opportunities for ongoing research on the issues discussed in this chapter. Supplementing that data, the Education Policy and Data Center, which was founded in 2004 and is a research unit within the Global Learning Group of FHI 360, now has a database that includes subnational statistics for over 60 developing countries (www.epdc.org). By increasing the amount of information readily accessible to researchers and practitioners in international education, data centers of this type can facilitate additional high-quality research on the topic of international equity in education.

Notes

1. The United States does not recognize education as a fundamental right and thus is not a signatory to these treaties. Education is not mentioned in the U.S. Constitution and the legal basis for promoting educational equity in the U.S. context is found in constitutional provisions at the state level that typically require the various states to provide a “uniform” or a “thorough and efficient” system of education.

2. For summaries of global assessments of student achievement, see UNESCO (2000b) and Kellaghan and Greaney (2001).

3. A good summary of these obstacles to gender equity can be found in Chapter III of From Access to Equality: Empowering Girls and Women through Literacy and Secondary Education (UNESCO, 2012).

References


Section IV
Changing Patterns of Governance and Finance

SECTION EDITOR: ANDREW RESHOVSKY
The Changing Federal Role in Education Finance and Governance

NORA E. GORDON

Introduction

This chapter details the transition from a historically nearly absent federal role in education finance and governance in the United States, through significant expansions in federal influence in the post–World War II period, to the current peak of federal involvement in educational policy and practice at the classroom level following the implementation of the No Child Left Behind Act of 2001 (NCLB) and the establishment of the Race to the Top Fund.

The U.S. Constitution does not give the federal government jurisdiction over education, and the 10th Amendment reserves all rights not explicitly granted to the federal government to the states. Thus the relatively (from a comparative international perspective) decentralized educational system in the United States is not surprising. The states historically allowed school districts to play the dominant role in the governance and finance of elementary and secondary education (see McGuire, Papke, & Reschovsky, this volume). In recent decades, however, the states have become significant contributors to education revenue—on average about equal to local districts, but with great variation across states (see Corcoran & Evans, this volume). A common and not inaccurate characterization of the federalist evolution of education finance and governance in the United States is that much power has shifted from local districts to states and that the federal role has been relatively small over time. A more complete characterization, however, would acknowledge significant expansions in the federal role in recent decades and the potential for still more to come.

The magnitude of the federal role cannot be measured fully using any one dimension. It may be quantified most simply by the amount of federal funds directed towards educational activities through the legislative process. Congress may affect educational practice more through the conditions it attaches to the receipt of federal funds than through appropriations alone, however. Because of constitutional limitations on federal powers, Congress relies primarily on such grant conditions rather than simply enacting laws that dictate behavior by state and local agencies. In principle, state governments and local school districts have the option of passing up federal funds, remaining unaffected by whatever conditions have been placed on their receipt.

Once Congress has set forth conditions for receiving federal grants, federal agencies such as the Department of Education write policy guidance that interprets the law in greater practical detail and has significant impact on the implementation process. They also influence policy in practice via less formal channels, such as how they advise state and local education agencies to prepare their applications for competitive grants or for waivers from existing policy requirements. Federal courts also play an important role in influencing the behavior of state and local educational agencies. The evolution of the federal role is marked by important changes by all these players, in all their decision-making capacities.

What Should the Federal Role Be?

While theoretical arguments can be made in favor of both a strong and a weak federal role, the practical political reality in the United States is that local governments, historically, have dominated the control and financing of education. There remains great resistance to reductions in this local control, and voters perceive...
local control to be closely linked to local finance. All this local control was at the discretion of the states, and current policy debates about how far federal education policy should reach centers around the tension between states’ rights and federal power.

Academic arguments for a relatively small federal role typically draw on the efficiency implications of the Tiebout (1956) model of local public finance, in which many small geographically distinct jurisdictions, such as school districts, provide public goods. The Tiebout model predicts that these jurisdictions and the bureaucrats who run them will compete against one another for residents, thus ensuring that residents receive high-quality, low-cost services. In this model, households with different tastes can choose the school districts that best meet their desired levels of local taxes and school spending. Hoxby (2000) finds that geographic areas with more school districts (and therefore more competition for residents) for reasons of geographic accident, such as more rivers and streams, produce higher test scores for lower educational spending levels, in support of these efficiency arguments and the Tiebout model. Rothstein (2007) critiques Hoxby’s work; see also her 2007 reply. Their detailed exchange has been followed closely because there has been so little empirical work, aside from the paper in question, using variation in centralization that is not likely to be correlated with variation in other determinants of efficiency in educational production; there is still less work finding statistically significant positive effects of jurisdictional competition on efficiency in the production of educational outcomes.

Arguments for a strong federal role, more akin to the pattern found in other industrialized nations, often emphasize the limited ability of state and local governments to redistribute resources within jurisdictional boundaries in a context of inequality across such boundaries. These arguments also rely on the fact that the quality of education in one school district can affect the well-being of residents of other districts through mechanisms such as worker productivity (Moretti, 2004b), crime (Lochner & Moretti, 2004), and civic participation (Dee, 2004; Moretti, Milligan, & Oreopoulos, 2004). Residents, therefore, may wish to subsidize education outside of their own districts, requiring a more centralized level of government to be involved. Finally, to the extent that some education-related activities have large fixed costs, such as research and development, centralized provision allows efficient pooling of resources to operate at scale.

**Growth of the Federal Role**

In practice, the federal role in American education often has grown for reasons not directly related to the previous economic arguments about how to optimally finance and govern an educational system, but rather for reasons more closely related to the political and economic issues of the day. These issues have ranged from concerns about national security to concerns about civil rights and the War on Poverty. As the following summary of key instances of major expansions in federal powers reveals, these noneducational issues were often instrumental in generating the political will necessary to expand the federal role. It is only relatively recently, beginning with the 1983 release of *A Nation at Risk*, a report issued by the National Commission on Excellence in Education at the request of the Reagan administration, that the federal push for greater involvement has focused exclusively on educational quality for its own sake. This coincides with a shift towards focusing on educational quality broadly, rather than attempting to improve conditions for particular categories of typically disadvantaged students, such as the poor or the disabled, through distinct categorical programs.

The remainder of this chapter first provides a brief and roughly chronological overview of milestones in the federal role in U.S. elementary and secondary education. It then reviews issues of current interest with respect to the federal role and lays out some of the major issues to be resolved in defining the ongoing federal role.

**The Evolution of the Federal Role in the United States**

The first federal education agency was established in 1867 with a commissioner and a staff of three (Office of Education library—early years, n.d.). This agency changed its name, bureaucratic status, and home several times before gaining the prestige of a cabinet-level agency, the Department of Education, in 1980. (See Goldin, 2006, for more on this bureaucratic evolution.) Throughout most of this period, the federal role in governing
and financing schools remained small. Congress did appropriate limited funds to states, but it did not interfere in decisions about curriculum, attendance requirements, graduation, or promotion (Goldin, 2006). When the federal government increased its role in educational governance, it tended to do so in conjunction with the award of federal revenue in the form of conditional grants-in-aid.

Changes in the amount of federal revenue to local educational agencies reflect many changes in the importance of the federal role over time. The earliest estimate of federal revenue to public elementary and secondary schools is from 1917, when it totaled approximately $1.7 million and comprised just 0.2 percent of total revenue received by local educational agencies, or school districts (Goldin, 2006). By 2009, federal revenue totaled about $76 billion and comprised 12.7 percent of total revenue (Digest of Education Statistics, 2012a). As Figure 19.1 shows, this transition was marked by periods of substantial increases in federal spending post-World War II and again in 1965, with the passage of the Elementary and Secondary Education Act (ESEA). A major theme in this evolution is the extent to which major federal policy changes closely follow social issues outside the education realm.\(^3\)

![Figure 19.1](image)

**Figure 19.1** Federal, State, and Local Shares of Primary-Secondary Revenue (selected years 1919–2009).


**Early Federal Education Policy: Support for Manufacturing and Agriculture**

The National Vocational Education (Smith-Hughes) Act of 1917, Public Law 347, marked the first major federal foray into elementary and secondary education policy and a substantial increase in federal funding.\(^4\) The Act resulted from concerted advocacy efforts by representatives from manufacturing, agriculture, and labor; it allocated $1.5 million (in current dollars) to agricultural and vocational education in 1918. This amount, comprising about 90 percent of total federal spending on education at that time, accounted for well under 1 percent of total revenues for elementary and secondary education. In contrast, as demonstrated in Table 19.1, in fiscal year (FY) 2013 the Department of Education’s total spending on elementary and secondary vocational education comprised only 3 percent of the Department’s total spending on elementary and secondary education.

Have federal vocational education funds been well spent? It is difficult to assess the success of vocational or
“career-technical” education (CTE) because students self-select into vocational coursework. The impact of the coursework itself on later academic and labor market outcomes is thus inextricably linked to the type of student who chooses that curriculum. The literature estimating effects of vocational coursework, therefore, attempts to control for as many student characteristics as possible to limit bias from unobservable student characteristics. Its findings are mixed. In explaining its initial budget proposals to eliminate vocational education funding completely, the Bush Administration cited the National Assessment of Vocational Education’s June 2004 Final Report that “found no evidence that high school vocational courses themselves contribute to academic achievement or college enrollment” (U.S. Department of Education, 2005). Bishop and Mane (2004) review much of this literature and, using National Education Longitudinal Study data for the high school class of 1992, estimate positive effects of high school vocational coursework on later employment and earnings.

Table 19.1 Composition of Department of Education Elementary and Secondary Expenditures (FY 2013)

<table>
<thead>
<tr>
<th>Program</th>
<th>In thousands of dollars</th>
<th>Share of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESEA Title I Grants to Local Educational Agencies</td>
<td>13,760,219</td>
<td>38.9%</td>
</tr>
<tr>
<td>Striving Readers</td>
<td>151,378</td>
<td>0.4%</td>
</tr>
<tr>
<td>Impact Aid</td>
<td>1,223,649</td>
<td>3.5%</td>
</tr>
<tr>
<td>Improving Teacher Quality State Grants</td>
<td>2,337,830</td>
<td>6.6%</td>
</tr>
<tr>
<td>Mathematics and Science Partnerships</td>
<td>141,902</td>
<td>0.4%</td>
</tr>
<tr>
<td>21st Century Community Learning Centers</td>
<td>1,091,564</td>
<td>3.1%</td>
</tr>
<tr>
<td>State Assessments</td>
<td>368,300</td>
<td>1.0%</td>
</tr>
<tr>
<td>Indian Education</td>
<td>123,939</td>
<td>0.4%</td>
</tr>
<tr>
<td>Teacher Incentive Fund</td>
<td>283,771</td>
<td>0.8%</td>
</tr>
<tr>
<td>English Language Acquisition</td>
<td>693,418</td>
<td>2.0%</td>
</tr>
<tr>
<td>Special Education</td>
<td>11,980,472</td>
<td>33.9%</td>
</tr>
<tr>
<td>Career and Technical Education</td>
<td>1,071,866</td>
<td>3.0%</td>
</tr>
<tr>
<td>Other</td>
<td>2,129,997</td>
<td>6.0%</td>
</tr>
<tr>
<td>Total, Elementary and Secondary</td>
<td>35,359,300</td>
<td>100.0%</td>
</tr>
</tbody>
</table>


Perhaps the most convincing evidence on vocational education comes from a randomized study of career academies by the nonprofit policy research organization MDRC (Kemple, 2004). They found that academy attendance had a significantly positive impact on labor market outcomes for males, but no labor market impact for females and no impact on educational attainment for either males or females. This provides strong evidence that well-targeted investments in vocational programs can improve job prospects, at least for men, without hurting educational attainment.

World War II: Impact Aid and the GI Bill

The federal government’s role in education began to grow in earnest during World War II. The Lanham Act of 1940 provided funds for home construction in communities with tax-exempt military plants and depots, and its 1941 reauthorization expanded the Act to include funds for school construction and operations (Kaestle, 2001). These efforts were extended with the 1950 impact aid laws for school construction and operations, and were later incorporated into the Elementary and Secondary Education Act under Title VIII. While current impact aid legislation still funds school operations and construction in federally impacted areas, the majority of its funding follows “federally connected children” such as members of Indian tribes to their respective school districts, which need not be impacted districts (About impact aid, n.d.). Table 19.1 shows that $1.2 billion was appropriated for impact aid in 2013, comprising about 3.5 percent of the Department of Education’s elementary and secondary education budget. Both the Lanham Act aid and the Impact Aid program were uncontroversial at their inceptions; political support for their beneficiaries was strong, and the programs provided funds to districts while explicitly retaining local control over the funds.
The best known federal education response to World War II was the GI Bill (the Servicemen’s Readjustment Act of 1944). It included, among other programs, assistance for postsecondary education and training of returning veterans and has continued to serve veterans through the present day. Although the main impetus for the original GI Bill likely was to prevent unemployment among returning veterans, the Bill had a substantial impact on educational attainment. Stanley (2003) finds that the World War II and Korean War GI Bills increased college attainment of the cohort most likely to be affected—men born between 1921 and 1933—by 15 to 20 percent.

The Cold War and Science Education

The launching of Sputnik in 1957, accompanied by widespread fear about U.S. ability to compete in math and science during the Cold War, generated the political momentum necessary for another categorical federal education bill in 1958, the National Defense Education Act (NDEA). The Department of Health, Education, and Welfare actually had produced a bill "nearly identical to NDEA" (Kaestle, 2001) in 1957 prior to Sputnik, reflecting a more general desire for the federal government to improve math and science education. Most of the federal funds allocated under NDEA were for higher education student loans and for math, science, and foreign language education at the elementary and secondary levels. A number of different and permanent federal programs since have absorbed the substantive components of NDEA (Clowse, 1981). Recent years have witnessed more public federal discussion of the importance of STEM (science, technology, engineering, and mathematics) fields, and the Department of Education lists a number of (fiscally minor) STEM-related programs, but there is not currently a single central program analogous to NDEA.

The Civil Rights Movement and Racial Desegregation

The federal role continued to expand not only through legislation but also through key court decisions, often driven by civil rights concerns. In Brown v. Board of Education (1954), the Supreme Court reversed its previous endorsement of the “separate but equal” doctrine of school segregation established by Plessy v. Ferguson in 1896 (see Patterson, 2001). Not until the Civil Rights Act of 1964, however, did Southern schools begin to desegregate in earnest. Title VI of the Civil Rights Act of 1964 specified that federal funds could be denied to public agencies not in compliance, and Title I of ESEA of 1965, which provided compensatory education funding to districts based largely on child poverty, then resulted in sufficient federal funds to make Title VI relevant. Many have speculated that without the creation of the Title I program, the Civil Rights Act would have been less effective in inducing school desegregation (Boozer, Krueger, & Wolkon, 1992; Clotfelter, 2001; Rosenberg, 1991). Cascio, Gordon, Lewis, and Reber (2010) show that school districts with more Title I funding at stake were more likely to quickly achieve the token level of desegregation necessary to receive their Title I grants than were demographically comparable districts eligible for smaller grants.

The ultimate impact of any federal policy depends on state and local responses to it, and one much-discussed local response to desegregation policies is “white flight” to residentially segregated school districts. Reber (2005) uses variation in the timing of implementation of major court-ordered desegregation plans to assess the effects of such plans on segregation and white enrollment. She concludes that court-ordered desegregation plans, largely implemented after 1968, did induce decreases in white enrollment, but on net reduced segregation substantially. Clotfelter (2001) finds that factors historically established as contributing to white flight—the exposure of white students to blacks in local public schools and the availability of alternative public school districts in the metropolitan area with lower black enrollment shares—continued to affect white enrollment in the 1980s and 1990s.

A robust body of evidence shows black students benefited from desegregation. Guryan (2004) finds that court-ordered desegregation plans account for about half of the drop in black high school dropout rates from 1970 to 1980, with no effect on white dropout behavior. Ashenfelter, Collins, and Yoon (2006) note that resources began to improve in black schools relative to white schools in the South well before Brown and find that state efforts to make black schools “equal,” at least in easily measurable and litigable dimensions (i.e., expenditures), had significant positive impacts on labor market outcomes for African Americans while schools
remained segregated. They also find that desegregation further improved these outcomes. Their findings on resources are consistent with Card and Krueger’s (1992) findings that increased education quality, as measured by teacher-student ratios, teacher salaries, and length of school year, in black schools relative to white schools in the segregated South prior to 1967 explains about 20 percent of the narrowing of the black-white male wage gap from 1960 to 1980.

Despite well-established benefits to desegregation, judicial decisions have lessened the federal pressure on school districts to fully desegregate in recent decades. One strand of decisions made it easier for districts with court-ordered desegregation plans to be declared “unitary” and released from the supervision of the courts: see Board of Education of Oklahoma City v. Dowell (1991), Freeman v. Pitts (1992), and Missouri v. Jenkins (1995). The net effect of these rulings has been to reduce the number of school districts supervised by the courts. This has been shown to increase school segregation (Lutz, 2011; Reardon et al., 2012) and, outside of the South, to increase both high school dropout rates and private school enrollment rates for black students (Lutz, 2011).

Following close to two decades marked by increased unitary declarations, the Supreme Court issued a landmark and complex decision in 2007 in Parents Involved in Community Schools v. Seattle School District #1 (PICS). As McDermott, DeBray, and Frankenberg (2012) describe, “the bottom line of PICS was unusually blurry” due to the range of opinions. The clearest aspect of the ruling is that school districts not under the court’s supervision are no longer permitted to use race as a “tie-breaker” in student assignment. The impact of PICS on student assignment policies over the long run remains to be seen.

Civil Rights for Language Minority Students

Congress first legislated how language minority students should be educated, and first allocated federal funds specifically for this group in the Bilingual Education Act of 1968. The Act created Title VII of the Elementary and Secondary Education Act, which awarded grants on a competitive basis to a small number of bilingual education programs, in which students are taught new subject matter content in their primary languages rather than in English. Title VII evolved over time to allow these funds to be targeted to the education of language minority students without restricting their use to bilingual programs. In NCLB, Title VII was eliminated as a separate program, consistent with the current federal emphasis on teaching English to language minority students. The magnitude of federal funding for programs for language minority students consistently has been a small share of federal education spending and has not grown commensurately with the language minority population of U.S. public schools. Table 19.1 shows that current federal funds for English language acquisition programs constitute only 2 percent of the Department of Education’s elementary and secondary budget. Just under 10 percent of all public school students in the United States received some type of services for English language learners in the 2010–11 school year, but with enormous geographic heterogeneity. For example, less than 1 percent of students in West Virginia received such services, while over one-quarter of California public school students did (Digest of Education Statistics, 2012b).

Title VII funded some programs directly and may have encouraged states to develop similar categorical programs of their own. Far greater federal impact, however, stemmed from the Supreme Court’s interpretation of the Civil Rights Act as applicable to language minority students in the 1974 Lau v. Nichols case. It required schools to “establish programs”—and specifically mentioned isolated bilingual classrooms as one acceptable type of such programs—for students who do not speak English so they may have “meaningful” participation in class, and ruled that failure to do so constituted a violation of the Civil Rights Act. As Nelson (2005) discusses, this requirement was particularly burdensome to districts also faced with a requirement that their schools be integrated if possible.

Since Lau, there has been considerable controversy over whether bilingual education, in which students are taught subject-specific material in their native languages, or English immersion, in which students are taught academic subjects in English while typically participating in additional English as a second language courses, is a superior method for teaching language minority students. Evidence to date on the effects of bilingual education on English reading and math achievement is mixed, with recent studies that aim to correct for known selection problems finding evidence of negative to neutral effects (Conger, 2010; Jepsen, 2010). Most of these studies focus on short-run effects of large-scale programs. In practice these programs may lack key elements expected to correspond to quality, such as strong training in teaching bilingual education. Rumberger
and Gándara (this volume) provide much greater detail on bilingual education, including state-level movements towards English immersion.

The War on Poverty: ESEA and Federal Approaches to Redistribution

The greatest single increase in federal funding for elementary and secondary education to date came with the passage of ESEA of 1965. The largest component of ESEA 1965 was Title I, designed to assist school districts in providing programs for poor, “educationally disadvantaged” children. As with earlier war-related efforts, the design of Title I reflected the more general social agenda of the times, the Johnson administration’s War on Poverty, in its focus on providing “compensatory education” funds to school districts based in large part on the number of poor children in the district. Title I dramatically affected federal school funding almost immediately. During the fall of 1965, Congress doubled federal expenditure on education when it appropriated almost $1 billion for the new program. For decades, Title I has remained the cornerstone of federal education policy. Title I ESEA remains the Department of Education’s largest single program, and as shown in Table 19.1, comprises nearly 40 percent of its expenditures. The 1994 reauthorization, the Improving America’s Schools Act, reflected the growing movement for standards-based reform, and the 2001 reauthorization, NCLB, brings the types of accountability provisions previously present in some states, but not at the federal level, to federal policy.

The formula for allocating Title I funds to school districts has changed over the years, but its basic components have not. School districts are allocated funds primarily based on the number and concentration of poor children who reside in the district and on the average level of school spending in the state. The program is not fully funded, which means that the amount appropriated for the program does not equal the sum of its grant obligations as determined by the formula. The current funding formula includes a “hold harmless” provision to prevent districts from experiencing declines in funding commensurate with any declines in the number of poor children, and also a “small-state minimum” which ensures that the sum of district allocations in each state meets some minimum level. Title I speaks directly to the redistributive motivation for a federal role. Given that most inequality across school districts comes from across rather than within states (see Corcoran and Evans, this volume), leaving school finance entirely to states and districts would severely limit the capacity for redistributive funding.

Title I targets funds to districts and schools by child poverty, while historically mandating that schools target funds to individual students whose academic performance is inadequate rather than on the basis of economic disadvantage. To the extent that the letter and spirit of the law prevailed, Title I maintained the federal tradition of targeting funds to specific categories of students. As lawsuits from advocacy groups reveal, Title I was in fact treated more like general aid in many cases in its early years. This resulted in increased legal requirements for districts to show how their use of Title I funds created supplemental opportunities for targeted students, as well as increased enforcement of compliance requirements, with districts required to pay back misused funds. These changes made Title I a less flexible resource for districts. Over time, Congress also has permitted more general uses of Title I funds, most notably by allowing schoolwide (as opposed to targeted assistance) programs in schools meeting various poverty thresholds over time. These schoolwide programs support efforts such as hiring more teachers to reduce class sizes.

The magnitude of the Title I program and its continued existence for half a century naturally prompts the question of whether it has improved educational quality and outcomes. This question proves quite difficult to answer. As attempts to evaluate the program have noted, there are critical selection problems given the targeting of Title I funds to schools serving poor students and to poorly achieving students within these schools (Puma, 1993). Another challenge in evaluating Title I is defining the treatment, given the potential for some share of Title I resources to supplant other (local and state) sources of support rather than increasing net inputs.

The extent to which federal funds supplement existing revenue unsurprisingly appears to depend on the context. Cascio, Gordon, and Reber (2013) investigate this question in Southern school districts in the late 1960s and show that when Title I funds were large relative to existing local revenue for education, Title I funds did serve almost entirely as supplemental. In these cases, Title I funding led to increased education spending and improvements in high school graduation rates for white students. When districts were eligible for Title I grants that were smaller relative to local revenue, however, significant shares of Title I were “crowded out” and there were no statistically significant effects on spending or graduation.

285
Studies using more recent data suggest Title I funds crowd out both the aggregate level of local revenue for schools and the distribution of local funds to Title I schools. Gordon (2004) investigates the possibility that Title I funds are used to substitute for rather than supplement state and local revenue. She finds that the revenue generated by local districts and given to districts by states changes in response to changes in Title I funds, so total revenue and school spending do not increase significantly with Title I grants. Consistent with these findings, van der Klaauw (2008) uses a regression discontinuity approach to compare New York City schools just eligible for Title I funds with their just-ineligible counterparts and finds that the two groups of schools do not have significantly different instructional spending—or student outcomes. Roza (2010) describes multiple districts in which categorical funds such as Title I crowd out unrestricted local resources at the school level.

**Civil Rights for Disabled Students**

The disabled were the next category of students to have their rights explicitly delineated. In 1975, Congress passed the Education for All Handicapped Children Act (P.L. 94–142) and President Ford signed it into law. It took effect in 1977, and was renamed the Individuals with Disabilities Education Act (IDEA) in its 1990 and subsequent reauthorizations. Unlike the *Lau* ruling for language minority students, IDEA does not simply mandate changes in district behavior. It provides funds for the education of students with physical and mental disabilities; IDEA is the second-largest single Department of Education program, constituting about a third of its expenditures (Table 19.1). Perhaps most importantly, it requires agencies receiving any federal education funds to provide a “free appropriate public education” in the “least restrictive environment” to all children, regardless of disability status. Just before the law’s passage in 1975, the Department of Education spent 3.7 percent of its elementary and secondary budget on special education. By 1980, this share had grown to 12.4 percent, and by 2013, to 33.9 percent. Federal funds, however, represent less than one-third of special education expenditures. The most recent national study of special education spending, conducted in 1999–2000 (Chambers, Parrish, & Harr, 2002), and data collected a few years later in a nine-district nationally representative sample (Alonso & Rothstein, 2010), both found that special education expenditures were 21 percent of total school spending.

The more critical aspect of the federal role in special education is likely to be the legal heft of the requirement for a free and appropriate education for all children, though there is considerable state-, district-, and school-level variation in how this is interpreted. In 2011–12, 12.9 percent of students in public schools nationally were classified as having a disability that rendered them eligible for special education services, reflecting a range from 8.8 percent of students in Texas to 17.5 percent in Massachusetts (Digest of Education Statistics, 2013). Hibel, Farkas, and Morgan (2010) use the Early Childhood Longitudinal Survey to reveal school-level variation in which students are identified for special education, conditional on individual characteristics, even within the same districts. For more on special education funding and policy, see Parrish, Harr-Robins, and Chambers (this volume).

**Federal Education Aid From Outside the Department of Education**

Given the tendency for federal education policy to come from other social, political, and economic pressures, it is not surprising that until 2003 more than half of spending by the federal government on elementary and secondary education came from agencies other than the Department of Education. In 2012, the Department of Education was responsible for about 59 percent of federal education spending, including all levels of education (New America Foundation, 2014). The Department of Agriculture spent more than any other agency through its support of school lunches and related programs. (See Schanzenbach, 2009 on school lunch and Bhattacharya, Currie, & Haider, 2006 on school breakfast programs.) Other sizeable federal programs include Head Start (under the jurisdiction of the Department of Health and Human Services), job training programs (Department of Labor), schools for children of military personnel (Department of Defense), and spending on Indian education (Department of the Interior).

The Head Start program is of particular interest in current policy discussions, in part because of the growing research base on the importance of fostering the development of noncognitive or “character” skills in early
childhood. It began in 1965 as part of the War on Poverty and provides early childhood education along with health and nutrition services for economically disadvantaged children, mostly ages three and four. Overall, the literature has identified positive but short-lived effects on academic achievement (for a review of earlier literature, see Karoly et al., 1998) and positive long-term effects on nonacademic outcomes. Garces, Thomas, and Currie (2002) compare outcomes across siblings with differential exposure to Head Start. They find that effects vary by the race of the student, with positive educational attainment and labor market outcomes for whites, and significant reductions in criminal activity for blacks. Using the same identification strategy, Deming (2009) identifies short-term cognitive effects that fade out quickly, but substantial positive long-term effects on educational attainment and health (mainly for females), and reductions in idleness (for males). In contrast to Garces, Thomas, and Currie, Deming identifies strong educational attainment impacts for blacks. Ludwig and Miller (2007) use a discontinuity design determining which counties received assistance for writing grants for Head Start, and find large and statistically significant positive health impacts, with reductions in child mortality due to diseases for which children can be tested and treated, but not due to other causes of death.

Head Start has never had sufficient funding to serve all eligible children (children in households below the federal poverty line). About half of eligible three- and four-year-old children currently are served by Head Start. In the context of shortages of slots in Head Start programs and the emerging research base on early childhood development, the National Head Start Impact Study (Puma et al., 2012)—the first randomized evaluation of the program—unsurprisingly has been the subject of widespread popular and academic attention. It studied children entering Head Start in fall of 2002; results to date follow those children through third grade. Overall, the study finds improvements in various measures for those randomly assigned to Head Start while they are in Head Start, but not once they move to elementary school, where they experience similar quality schooling as those assigned to the control group. These subjects are not yet old enough to use in a study of long-term outcomes.

A Nation at Risk and the Charlottesville Summit: Moving Toward a Stronger Federal Role

In a departure from the civil rights and antipoverty motivation driving much federal education policy in the 1960s and 1970s, the National Commission on Excellence in Education’s 1983 report A Nation at Risk reflected concern with the state of America’s schools from the perspective of their impact on national competitiveness. The first two sentences of the report make this clear: “Our Nation is at risk. Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world” (National Commission on Excellence in Education, 1983, p. 5). Through use of the bully pulpit rather than any new funding or legislation, A Nation at Risk prompted what Fiske (1991) describes as “an era of feverish educational reform efforts” in the 1980s. While President Reagan’s then Secretary of Education, T. H. Bell, created the Commission and the Administration released the report, the reform efforts following it were largely at the state rather than federal level.

President George H. W. Bush, who had run for office as “the education president,” began his term by convening a meeting of the nation’s governors to discuss education policy. At this meeting, at the University of Virginia in Charlottesville in September of 1989, attendees set goals for the country’s schools to meet by 2000. With a new Secretary of Education, Lamar Alexander, in place, President Bush released his America 2000 plan for school reform in April 1991. While A Nation at Risk pointed out deficiencies in American schools, America 2000 made specific proposals to address these problems. These included school-based innovation (through New American Schools, a not-for-profit effort to implement and evaluate promising comprehensive school reform models); national standards and tests; school report cards; and some federally supported school vouchers. Although most input from the nation’s governors in the 1989 education summit did not translate into immediate changes in federal education legislation, the summit was critical in establishing the role of the states as advisors to the federal stance on education, and many components of the recommended changes returned to the policy table in subsequent years.
The Federal Role in the Twenty-First Century

With its 2001 enactment of NCLB, Congress reauthorized and fundamentally changed ESEA, increasing the scope of federal powers by using the carrot of Title I funds to induce states to adopt test-based accountability systems. Though NCLB was modeled on systems already in place in a number of states, it was widely viewed as a significant expansion of federal power over states and their schools. If political observers at the time had known what was to come in the next decade, however, they may well have viewed NCLB as a relatively moderate policy change. Drawing on additional resources from the American Recovery and Reinvestment Act (ARRA), as well as increased executive branch power from the Department of Education’s NCLB waiver process, the federal government followed its focus on accountability with NCLB by expanding its powers related to teacher evaluation and the content of standards—topics perhaps even more controversial than the principle of test-based accountability. This has resulted in a federal role that is not only historically unprecedented in its strength and scope, but also extremely politically contentious.

No Child Left Behind

NCLB reauthorized the funding stream for ESEA, adding significant accountability provisions and other conditions to federal aid. Title I schools are monitored for their adequate yearly progress (AYP) in working towards having all students and all sufficiently large “subgroups” of students (determined at the state level) meet state-determined proficiency levels by 2014. Various penalties exist in the law for failing to meet AYP, with the severity of the penalty depending on the number of consecutive years of failing to meet the standards. The most drastic penalties involve restructuring, with potentially significant changes in school governance. During the Obama Administration and particularly with the support of the ARRA, federal School Improvement Grants have provided resources for a subset of schools facing this consequence and required schools to implement one of four prescriptive interventions as a condition of receiving these funds.

Under NCLB, states retain authority over the content of their standards, their testing instruments, and the cutoff scores used to determine proficiency on their tests. They also control a number of less politically visible decisions. For example, they choose confidence interval adjustments and minimum subgroup size; Davidson et al. (2013) show that these dimensions afford states significant power in determining the share of schools making AYP, conditional on student performance.

The law was due for reauthorization in 2007 and has been funded since then via continuing resolutions. During this time, it has been clear it would be enormously disruptive to continue implementing the unamended law. For example, in 2010–11, 48 percent of schools were estimated as failing to make AYP (Usher, 2012). In acknowledgment of this reality, the Department of Education established a waiver process by which state and local education agencies receive “flexibility” from some NCLB requirements in exchange for state improvement plans (U.S. Department of Education). Since the waivers were announced in September 2011, 42 states, the District of Columbia, and Puerto Rico have been granted waivers (U.S. Department of Education, n.d.a). The importance of the waiver process in practice has been to shift education policy decision making from the legislative to the executive branch. For example, states must demonstrate their college and career ready standards, and adopting the Common Core State Standards—notably not a condition of aid under the law—provides states with a speedy way to do so.

There is strong consensus that NCLB sparked significant changes in educational practice; note all studies described here predate the flexibility era. Jennings and Rentner (2006) describe schools increasing time spent teaching reading and math, aligning their curricula to better match state standards and tests, and increasing attention to the subgroups of students most likely to cause schools to fail to meet AYP goals. Reback, Rockoff, and Schwarz (forthcoming) use data from the Schools and Staffing Survey to examine effects of NCLB on teachers’ attitudes and hours, focusing on teachers in schools likely to be feeling the greatest pressure from NCLB due to their risk of failing to meet AYP. They find that teachers in these schools express greater concerns over the role of student achievement than in their job security. Untenured teachers at these schools also report working longer hours. Dee, Jacob, and Schwartz (2013) find NCLB led to an increase in teacher compensation and an increase in the share of teachers with graduate degrees.
These changes in practice appear linked to improvements in student learning. Dee and Jacob (2011) exploit the fact that many states were already operating accountability systems prior to NCLB, and compare changes in scores from the National Assessment of Educational Progress across states experiencing larger and smaller changes in their policy environment due to NCLB. They find that NCLB has led to large improvements in fourth-grade math scores, smaller but still statistically significant improvements in eighth-grade math scores, and to no changes in reading scores. Reback, Rockoff, and Schwarz (forthcoming) examine changes in student achievement on the low-stakes test administered in conjunction with the Early Childhood Longitudinal Survey. They identify modest (.07 of a standard deviation) positive effects on reading scores and no statistically significant effect on math scores in schools “under pressure” from NCLB. They also find that NCLB pressure does not appear to affect students’ self-reported enjoyment of learning or their anxiety related to testing.

ESEA has been the dominant federal elementary and secondary education funding mechanism for the past half century. At present, political challenges make the reauthorization process appear intractable. Combined with the discretionary nature of ESEA flexibility, it is difficult to speculate on the future of the law.

**Race to the Top**

While the Obama administration may have found a back door for policymaking via ESEA flexibility, it has pursued many of the same policy goals openly through the competitive Race to the Top (RTT) Fund program. RTT has awarded over $4 billion to date to states implementing reforms and innovations focused on student achievement gains. While RTT garnered much media attention, it is important to note that it is not nearly as fiscally significant as Title I ESEA, which distributed over $14 billion per year in recent years. In the first three rounds of Race to the Top in 2010 and 2011, states competed for federal funds; in 2012, the Department of Education offered a district-level version of the program. State and district applications were scored according to a rubric defined by the Department, with emphasis on improving student achievement as well as on adopting specific and controversial policies (e.g., establishing longitudinal student data systems, using student achievement growth to inform human resource decisions, and adopting common standards).

The expected benefits of submitting a competitive application—including statements representing cooperation from key stakeholders such as teachers’ unions and state legislatures—have persuaded states to enact policies they had previously resisted, even though doing so does not guarantee winning the federal funds. For example, in January of 2010, California enacted laws to implement longitudinal student-teacher data systems and open enrollment policies; ultimately its RTT application was not successful (Hudson & Keller, 2010).

As of January 2014, 18 states and the District of Columbia and 16 districts have “won” rounds of the Race to the Top (U.S. Department of Education, n.d.b). An additional program, the Race to the Top–Early Learning Challenge was created as a joint venture between the Departments of Education and Health and Human Services to provide grants to states investing in public education for low-income students between birth and age five (U.S. Department of Education, 2012). Between 2011 and 2013, 20 states were awarded Early Learning grants. The first evaluation of RTT commissioned by the Department of Education is due out in late 2014. While it is too early to note RTT’s impact on students, it is already clear that it has persuaded states to change their policies.

**Common Core State Standards**

The Common Core State Standards are a set of expectations for student performance, in mathematics and English language arts, in kindergarten through 12th grade. The National Governors Association and the Council of Chief State School Officers led the effort to develop the standards. Initially, all but five states (Alaska, Minnesota, Nebraska, Texas, and Virginia) adopted the Common Core. While the federal government neither determined the content of the standards, nor required states to adopt them, it certainly encouraged states to do so. For example, the NCLB waiver process (U.S. Department of Education, n.d.a) requires that states demonstrate their college and career ready standards. Race to the Top awarded 70 out of 500 points on a state’s grant application for working with other states to develop common standards and assessments, though
Options for the Federal Role Moving Forward

Much of the shape of federal education policy in the near future will be determined by the nature of ESEA in practice. ESEA will celebrate its 50th anniversary in 2015 and has been overdue for reauthorization since 2007. Will Congress come together sufficiently to play a role in federal education policy, or will the power remain with the executive branch through the waiver process? And if policymakers are unhappy with the practical implementation of ESEA but unable to change it, will they shrink it, perhaps while growing other programs?

Federal education policy is also about to face considerable new challenges. However much NCLB revealed inequality in educational outcomes across states, when states start using common assessments aligned with Common Core State Standards, they are likely to expose even more. This could prompt questions about whether all schools are equipped to meet the standards assigned to them and whether a greater federal role in financing is needed to ensure that they are. Corcoran and Evans (this volume) estimate that in 2011, 78 percent of the total variance in spending per pupil came from across rather than within the states. This is a historic high, and suggests that now more than ever state-level redistributive policies alone will be insufficient in moving towards greater equalization of resources.

If the federal government is to assume a greater role in redistributing resources to districts, it could do so either through a new general, foundation-type program or by expanding its revenue to districts through categorical programs. Though the federal government historically has not distributed aid through a program as general as a foundation aid system, such as the one Rothstein (2001) proposes, there have long been unsuccessful movements for one (Kaestle, 2001). The existence of many federal funding streams, each with its own devoted constituencies, may limit the ability of the federal government to move to a more streamlined and flexible approach to funding education and necessitate continued reliance on categorical programs. Ladd and Hansen (1999) review these two possible approaches in detail, while Kaestle (2001) cautions against viewing them as distinct when implementation may render categorical spending general in practice, particularly over time. Policies of the past decade suggest that whether funds are general or categorical, now more than ever the strings attached to those funds may be the defining characteristic of federal education programs.

Notes

1. McDermott and Jensen (2005) provide several examples of policies outside of the education realm which were struck down by the courts when implemented in a regulatory framework, but then ultimately prevailed when instituted as conditions for the receipt of grants, suggesting that this strategy increases the power of federal agencies over the states more broadly.

2. Education exhibits more private good qualities than public good ones, but the general argument of the Tiebout model holds.

3. Kaestle (2001) provides a useful discussion of how this historical evolution is often framed as “episodic” in its responsiveness to other social issues, but also reveals how in many cases episodic shifts have followed periods in which support for the agenda at hand grew incrementally, as well as examples of crises that did not ultimately change education policy and policies that changed without any precipitating external event. Kaestle refers the reader to Nelson Polsby’s (1984) work on “acute” versus “incubated” policies.

4. The first major federal education legislation was the Morrill Land Grant Act of 1862, in which Congress gave federal land to the states (those remaining in the Union during the Civil War) in proportion to their populations for the establishment of agricultural and technical colleges. The Second Morrill Land Grant Act, in 1890, appropriated land to the Southern states, and also appropriated federal funds from sales of public lands to the states for land grant colleges. New conditions were added with the new funds: the states could not use them to support colleges "where a distinction of race or color is made in the admission of students," but did allow this if the state maintained separate colleges for different races and divided the federal funds equitably among those colleges. This condition prompted the establishment of what are now known as "historically black colleges" (Goldin, 2006), and provides an early example of conditional grants, which emerge as a favored mode of federal education policymaking. The Morrill Land Grant Acts have had a lasting legacy on the geographic landscape of American higher
education. Their impact is so significant that Moretti (2004a) is able to use the original grants as an instrument for the city-level distribution of educational attainment in 1990. He finds that having a land-grant college in a city is associated with a 20 percent increase in the share of college-educated adults. The Smith-Hughes Act of 1917, however, was the most significant federal legislation applicable to elementary or secondary education at that time.

5. Career academies are public high schools with academic and vocational content organized around particular professions and linked to relevant local employers. The findings therefore should be interpreted as relevant to career academies but not necessarily to other types of vocational education.


7. These changes in Title I grants, however, are much smaller than the full grant amount received by any district, so it is not possible to extrapolate from these estimates to predict district spending levels in the complete absence of the program.

8. For example, Title I funds under ESEA are nominally categorical funds in that they are based on district-level child poverty and are to be spent on a particular program, “compensatory education.” Over time, Title I has turned into a program present in over half of public schools; with high-poverty schools now allowed to use Title I funds for schoolwide programs, including class size reduction, Title I funds are essentially general revenue for many schools.

References


The Role of Nongovernmental Organizations in Financing Public Schools

JANET S. HANSEN, MICHELLE HALL, DOMINIC J. BREWER, AND JANE HANNAWAY

Introduction

Financing K–12 public schools in the United States is overwhelmingly a governmental responsibility. The National Center for Education Statistics reports that in 2009–10, sources other than federal, state, and local governments accounted for just 2.0 percent of all the revenues received by public schools. Direct contributions to public education from outside sources have been on a gradual decline from a high in 1994 of 2.7 percent. Even this figure overstates the financial contribution of nongovernmental organizations that provide funds to support or improve public schools, since it includes income from various fees as well as contributions and donations from private sources.

Despite the small share of public school revenues provided by nongovernmental organizations, this funding source is of increasing interest from a policy perspective. There is a growing perception that tight school budgets are leading to a situation in which “public school coffers require a private boost” (James, 2011) through voluntary donations. Researchers have suggested that changes in public school finance systems (especially the growing dependence on state rather than local revenues) have encouraged districts and schools to turn to private fund-raising. Nongovernmental funding helps them to overcome state-imposed constraints on local spending and to meet new demands for improved educational performance. While private funds supporting public education are nothing new, anecdotes and some research evidence suggest that districts and schools “are now pursuing private support with increased sophistication and aggressiveness” (Zimmer et al., 2003, p. 485). Foundation activities in education, most recently that of the well-funded Bill & Melinda Gates Foundation, garner widespread public attention.

Educators appreciate nongovernmental revenues. These funds often come with fewer restrictions than do public dollars, and they support activities and programs that would otherwise be unaffordable. At the same time, a growing reliance on private donors raises a number of issues. Chief among them is whether nonpublic funding negates some of the equity gains that school finance system reformers have achieved in the distribution of public funds. There are questions about whether private funding offered for specific purposes inappropriately affects the decisions of public officials charged with making resource allocation and other strategic decisions for schools. Concerns have been raised about whether private donors, especially large foundations, are sufficiently attentive to the effectiveness, impact, and sustainability of their spending as well as the appropriateness of the extent of their influence on public education reforms.

This chapter provides an overview of nongovernmental funding of public schools in the United States. It begins by identifying the wide variety of organizations that contribute to public education and the kinds of private giving. An explanation of the data limitations plaguing this aspect of education will serve to indicate why it is difficult to say with any precision exactly how much nongovernmental revenue is directed at public schools and for what purposes. Whatever the amount, however, nongovernmental organizations support education in ways that range from enhancing traditional activities to fostering comparatively ambitious reforms and do so for reasons ranging from the purely altruistic to the largely self-interested. Thus, the remainder of the chapter describes the major nongovernmental organizations, the roles that they play, and the potential impact their spending has on educational policy and practice.
Nongovernmental Organizations That Finance Public Schools

There is no accepted typology of nongovernmental organizations that provide revenues to or on behalf of public schools. Nor are there common reporting arrangements that would allow an accurate and complete picture of this universe. A reasonable list of organizations (many of which will be defined further in subsequent sections) that make financial contributions would include:

- School-based organizations (parent associations, alumni associations, booster clubs)
- School foundations
- Local education funds
- Community foundations
- Local businesses
- Private foundations (including independent, operating, and corporate foundations)
- Corporations (via direct giving, business/school partnerships, and commercial relationships)

Some of these organizations do not limit themselves to financial contributions. For some the value of volunteer time or other in-kind contributions, for example, may outweigh their monetary assistance. Some organizations that infrequently make financial contributions to schools are not on this list; for example, colleges sometimes give money as well as other services to local schools (e.g., Nichols, 2006). The list emphasizes one type of “nontraditional revenue source”—contributions from donors (Addonizio, 2001)—over enterprise activities such as user and developer fees. The list does, however, note one kind of enterprise activity—commercial relationships—because this activity more than others reaches inside the school building and even inside the classroom. We also focus on organizations that provide some financial resources to schools and generally ignore other kinds of contributions, such as foundation support of education-related research at universities and think tanks, that may have as much or more impact on schools than do direct contributions. While acknowledging that the boundaries are fuzzy and that the organizations and activities sometimes overlap, the chapter examines nongovernmental organization financing provided to or on behalf of K–12 education by three distinguishable groups of donors: (1) locally based voluntary contributors, (2) private foundations, and (3) corporate supporters.

What Donors Give: The Limits of Data

There is no single source of data on the number of nongovernmental organizations that provide revenue to public schools or on the amount that they give. In overall terms, private giving is at best a very small proportion of the $597 billion in total revenues devoted annually to public elementary and secondary education. While we provide some evidence in the sections below about magnitudes of, and changes in, levels of private giving, we do not attempt to develop a credible estimate of totals. Researchers who have examined various nongovernmental funders have chronicled some of the difficulties in painting a statistical portrait of these organizations and their donations. As Jenkins and McAdams (2005) reported in their study of philanthropic giving in just three school districts, "Preparing a coherent account of this topic was … less like telling a story and more like assembling a puzzle from widely scattered pieces" (p. 134).

National statistics on district revenues typically draw on data collected by the U.S. Census Bureau in its "Annual Survey of Local Government Finances: School Systems." This survey has traditionally collected information on contributions and donations from private sources in a single reporting line that includes rentals, gains and losses on capital assets, refunds of prior year’s expenditures, and miscellaneous local revenues.\(^3\) Beginning with its Fiscal 2006 survey, however, the U.S. Census Bureau required separate reporting of contributions and donations from private sources. Distinguishing these funds from other miscellaneous sources of local revenue should provide better information on the private giving that is included in the financial recordkeeping systems of local districts. States and local districts may have more finely grained reports, but definitions and categories are inconsistent across sites and are difficult to interpret on their face.

Efforts to quantify giving by studying donor statistics run into similar problems of incompleteness and inconsistency. Many donors channel their contributions to public education through nonprofit entities...
established under section 501(c)(3) of the Internal Revenue Code. As of 2010, such organizations with gross receipts greater than $50,000 (the minimum amount prior to 2010 was $25,000) must annually file a Form 990 with the Internal Revenue Service, and some researchers have used these reports to try to develop a portrait of education giving (see, e.g., Brunner & Sonstelie, 1997; Brunner & Imazeki, 2005; Lampkin & Stern, 2003; Nelson & Gazely, 2014). However, many local contributors fall under the $50,000 threshold. For those above it, it can still be difficult to consistently identify those whose purpose is support of public education. Studies based on data that charities must report to individual states can surmount some, but not all, of the problems in pinning down giving that is targeted on public education. (See, e.g., California, as used by Brunner & Sonstelie, 1997, and Brunner & Imazeki, 2005; nationwide, as used by Nelson & Gazely, 2014; and Michigan, as used by Addonizio, 2000.)

Despite the effort of foundations to reduce the ambiguity over the purposes and amounts of their giving by voluntarily providing data to the Foundation Center, the most comprehensive database on U.S. philanthropy, uncertainty still abounds. In a study attempting to quantify philanthropic support of public K–12 education, Greene (2005) (who came up with an estimate of $1.5 billion annually) describes the imprecision of the Center’s education-reporting categories and inconsistent reporting across classifications by the foundations themselves.

Locally Based Voluntary Contributors

Local individuals and organizations have long provided financial support to public schools. What is new is the growing number of such contributors and the increasingly formal and sophisticated way in which many are approaching their activities.

A Picture of Local Contributors

While individuals and local businesses sometimes provide money, equipment, and other gifts directly to schools on an ad-hoc basis, private giving is increasingly channeled through a variety of organizations with missions that explicitly include fundraising for schools and/or districts.

School-Based Parent, Alumni, and Booster Clubs. These familiar organizations may bring in just a few hundred dollars; others are raising tens of thousands and using the money for such things as athletic tracks, computers with wireless Internet access, a new phonics curriculum, or salaries of instructors and teacher aides (Luttrell, n.d.).

School Foundations. Some districts and schools have moved to make local fundraising more formal and sophisticated by establishing foundations (tax-exempt 501(c)(3) organizations set up to support individual schools and/or whole districts). School foundations (as distinct from local education funds described below) were generally created to raise extra dollars for schools. They do not attempt to operate independently from the schools they support and may have staff appointed by the school or district and/or have school officials sitting on or chairing their boards (Lampkin & Stern, 2003).

According to the American Schools Foundation Alliance, there are approximately 7,000 public K–12 school foundations. The precise number of school foundations is difficult to pinpoint, in part because the IRS requires only foundations with revenues above $50,000 to file a Form 990 and many school foundations do not bring in that level of revenue in a given year. A handful, however, are quite large as measured either by absolute dollars or as a proportion of district or school spending. For example, the Hillsborough (CA) Schools Foundation reports that it gave $3.45 million to the district schools in 2013–14 school year, increasing the district budget by 17 percent. In New York City, the Fund for Public Schools received over $47 million in donations and pledges from fiscal year 2013.

Statewide and national associations exist to support the formation and programs of these foundations. The California Consortium of Education Foundations assisted 600+ foundations that raised more than $250 million for California school districts in 2009. The Oklahoma Foundation for Excellence supports the work of 195 foundations. There are also national associations that provide resources to help schools and school districts that want to establish foundations or improve and expand their financial performance including the National
School Foundation Association and the American Schools Foundation Alliance.\textsuperscript{11}

\textit{Local Education Funds.} While the boundaries are not always neat (and similar names may be used for different kinds of organizations), activists and researchers distinguish local education funds (LEFs) from school foundations on a number of dimensions. LEFs are also nonprofit organizations that support public schools, but they operate on a district, regional, or statewide level and act independently of the schools and districts themselves. They often focus on areas with significant numbers of low-income families. While school foundations use their resources mainly to support school programs, many LEFs see system-wide reform efforts as a key part of their missions. LEFs are on average larger than school foundations in terms of both revenues and expenditures and have larger paid staffs (Lampkin & Stern, 2003). They sponsor a variety of activities aimed at creating a supportive civic environment for public education; enhancing school and community capacity to support student achievement; and raising community status as measured by educational attainment, higher education participation, and economic development.

\textit{Community Foundations.} A final type of local contributor to public education is the community foundation. Legally distinct from private foundations—such as the independent and corporate foundations described below—a community foundation operates a grant program addressing broad public needs of the geographic community or region in which it is located. There were 734 such foundations in 2011 providing $4.1 billion in grants across all areas. The Foundation Center (2014b) reported in 2010 that 26 percent of all community foundation giving went toward education.\textsuperscript{12}

\section*{Concerns Arising From the Growth in Local Voluntary Contributions}

Even though local voluntary contributions represent a small fraction of expenditures on public education, they can be sizeable in specific circumstances. Thus, a number of concerns about these organizations have been raised, chief among them the possibility that growing reliance on local contributions would undo some of the effects of school finance reforms meant to reduce spending disparities among districts and schools.

\textit{Equity.} School finance reform, court rulings, and tax limitation laws in California in the 1970s had the effect of centralizing finance decisions at the state level, limiting what local officials could spend on public education, and reducing spending disparities among districts. These changes spurred growth in organizations to channel private contributions to schools and stimulated research aimed at assessing the extent and effects of this private giving. Much of what is known about local voluntary contributions to education comes from this California-based research. According to the National Center for Education Statistics, in 2001 there were 320 educational foundations, 1,463 parent-teacher associations, and 335 other organizations that filed Form 990s and raised over $238 million for California public schools.

Brunner and Imazeki (2005) found that, although private contributions amounted to just $39 per student statewide in 2001, they were not uniformly distributed. Schools with nonprofit organizations that raised $25,000 or more tended to be concentrated in high-income districts where school spending was most constrained by tax limits imposed by school finance reform. These districts had about three times ($135) the average private contribution per student than the state average. They also reported that the 2001 private contributions in the state were about twice the voluntary contributions in 1992 (inflation-adjusted) when public resources were less constrained. On average, the contributions did not lead to large resource inequalities across districts, though the impact of private contributions in the smallest and wealthiest jurisdictions was most pronounced.

Zimmer et al. (2003) also found in a sample of districts in Los Angeles County that while poorer communities have fewer monetary contributions from parents, they may be more apt to benefit from donations from organizations that concentrated on support for less-advantaged communities, including the philanthropic and corporate donors to be described below. These analyses are largely echoed by research on education foundations in Michigan (Addonizio, 2000).

Nevertheless, some districts have adopted policies to guard against the potential dis-equalizing effects of voluntary giving, especially among individual schools in the district. In Montgomery County, Maryland, the school board in 1989 prohibited outside groups from paying for teachers’ salaries. Later, the board specified that some kinds of facilities-related costs, such as stadium lighting and landscaping, can, with board approval, be paid for with private funds. A new gymnasium, however, cannot be privately funded unless it will be
available to the community at large. In Portland, Oregon, outside groups that raise more than $5,000 for
teachers’ salaries must send one-third of the money to the Portland Schools Foundation to use at schools in
low-income neighborhoods (Galley, 2003). According to the Center for Investigative Reporting (2012), the Santa
Monica-Malibu, Manhattan Beach, and Palo Alto school districts require the pooling of volunteer schoolwide
contributions at the district level for redistribution (as cited by Nelson & Gazely, 2014).

Sustainability. Questions arise about whether voluntary contributions can be sustained, especially if they are
used to pay for schools’ routine operating expenses. There are uncertainties about whether volunteer
fundraisers can keep the money flowing in year after year. Moreover, the leadership of local groups such as
parents’ organizations and booster clubs may change frequently, with new leaders having new priorities for
their funds. The concern about sustainability is one factor that motivates some local individuals to form
education foundations to act as full-time fundraisers and watchdogs over spending. It also has encouraged
some of these foundations to focus on setting up endowments to sustain beneficiaries in the future (Galley,
2003). As suggested earlier, some school boards address the sustainability issue by limiting private giving to
one-time expenditures and “extras” rather than paying for items that should be covered by routine operations.
Some school supporters also argue that private givers need to focus their activity not just on fundraising but on
building community support for public education more generally and for public funds to address important
school needs (Luttrell, n.d.).

Private Independent Foundations

Private foundations have been providing support for public education since early in the 20th century. As
foundation donors increasingly turn from ad-hoc project support to “reform-oriented giving,” their activities
arguably warrant more outside scrutiny.

Independent Foundations and Public Education

The foundations discussed in this section are the independent foundations; one of three types of private
foundations as defined by the Foundation Center. A second, the corporate foundation, is discussed in the next
section. The third, the operating foundation, is not relevant to this examination because operating foundations
conduct their own charitable, educational, or other exempt programs and activities outside the scope of public
school grant provision.

Independent foundations are the most prevalent type of private foundation, comprising 89 percent of those
in the Foundation Center’s database. An individual or family usually provides these foundations’ assets in the
form of gifts or bequests that are held as an endowment. Because of the narrow base of their support, they are
subject to the private foundation laws intended to ensure that they serve the public good.

There were about 73,764 independent foundations in 2011. The top 50 independent foundations (ranked by
total giving for all purposes) account for about a third of all independent foundation giving. The best
available estimate of foundation giving from the top 1,000 foundations in 2011 (from foundations of all types)
to public elementary and secondary education was approximately $1.8 billion. The disparity in giving levels
among foundations is apparent within the top 50 givers: the biggest K–12 donor (Bill & Melinda Gates
Foundation) gave $447 million in 2011; No. 50 (Marian Community Foundation) gave just over $17 million.
Not all this K–12 giving goes directly to public schools or organizations that directly support public education
activities. Independent foundations today are funding a broader array of education activities including
research, advocacy, human capital development, efforts to improve public school accountability and
wraparound service support programs.

The Track Record of Independent Foundation Philanthropy

Independent foundations have a long history of supporting public education, and their efforts have been so
diverse that summarizing them is impossible. As Frumkin (2005) notes:

One of the oldest and most popular targets for philanthropy is education. Going all the way back to the start of modern, large-scale giving, donors have found the idea of supporting education—in all its many forms—attractive because it offers one of the clearest and most compelling ways to increase opportunities.

(p. 275)

An early example was the General Education Board (GEB), established by John D. Rockefeller in 1902 with an initial endowment of $33 million. The largest of a number of endeavors created after the Civil War to foster education for black students, the GEB supported teacher training, rural school agents, and the development of state education departments and sponsored studies, demonstrations, and conferences.

Foundation-sponsored efforts have sometimes left a permanent mark on public schools. Colvin (2005) points out that "many of the features of public schools that are taken for granted can be traced to foundation-funded projects" (p. 25). He cites as examples classroom aides, the Advanced Placement Program, the creation of schools and colleges of education, the Educational Testing Service and achievement testing programs, middle schools as a replacement for junior highs, the academic standards movement, and the research and advocacy projects that spawned lawsuits across the country challenging the constitutionality of disparities in spending and tax rates between rich and poor communities.

Nevertheless, there is a belief among both researchers and foundation officials themselves that the result of a century of foundation philanthropy has been disappointing. Loveless (2005) states the argument bluntly: "The record of philanthropic foundations in promoting education reform is known more for its failures than its successes" (p. 105). His view is echoed by many of the other researchers and practitioners who undertook studies for Frederick Hess’s pioneering volume on K–12 philanthropy (Hess, 2005) and by Fleishman (2007) in his book on American foundations. They do not quarrel with the proposition that many individual projects positively affected their beneficiaries. However, they argue that foundations have not had a significant impact on improving the overall quality of American education.

Two major programs, separated in time by more than 30 years, symbolize for many the disconnect between the money and time invested and the results achieved in too many foundation undertakings.

In 1960, the Ford Foundation launched the Comprehensive School Improvement Program (described in Manno & Barry, 2001). It was not a "one-off" effort, but part of a larger systemic reform agenda with "a distinctive, well-articulated reform vision" aimed at social change. The foundation spent $30 million throughout the 1960s on 25 projects in school systems (not individual schools) encouraging a variety of innovations designed "to create a critical mass—a chain reaction of change that would overcome the inertia of school systems and produce significantly different educational institutions" (quoted in Manno & Barry, 2001, p. 2). The foundation undertook ongoing evaluation of the projects, cross-site conferences, and frequent site visits. In its 1972 report A Foundation Goes to School, the foundation itself (as reported by Manno & Barry 2001, pp. 3–4) documented the failure of CSIP to spur significant change because its sponsors failed to grasp the complexity of improving schools.

In the 1990s, the largest philanthropic grant ever made to American schools, the Annenberg Challenge, had similarly large ambitions but was ultimately also judged to have been generally ineffective. Walter Annenberg’s $500 million gift to nine large city school systems, a consortium of rural schools, two national school-based-reform groups, and several arts education projects were viewed in the end as supporting more of what was already going on in schools, rather than fostering major change (Colvin, 2005; Fleishman, 2007). Colvin reports varied explanations for the project’s lack of impact, including among others: not enough foundation staff to guide such a big effort; money spread too thin; activities focused too much on the margin instead of radically redesigned core school operations; and insufficient attention to systemic shortcomings such as adequate, equitable, and reliable funding of schools.

Producing lasting improvement in elementary and secondary education has proved difficult. As a result, some major foundation donors have withdrawn from the field, such as the Edna McConnell Clark Foundation, Pew Charitable Trusts, Rockefeller Foundation, and the David and Lucile Packard Foundation (Hess, 2005; Manno & Barry, 2001). While these foundations have decided to focus their funding elsewhere, other large-scale foundations, such as the Broad, Gates, and Walton Family Foundations, are using new strategies to invest in education (Reckhow, 2013).

Indeed, as we discuss below, recent foundation-funded initiatives suggest a narrowing of focus for
foundation grant making, evidenced by coordinated donations, fewer grant-ees, and larger total amounts. For example, recipients of the largest total K–12 grants from multiple foundations in 2011 included: Teach for America with over $73 million, the Charter Fund with over $47 million, and the Knowledge Is Power Program with over $22 million in donations. These grantee organizations operate across areas larger than the traditional single school district structure. Indeed, they implicitly (and sometimes explicitly) challenge traditional ways of doing school business across districts and states.20

**Recent Developments in Foundation Philanthropy**

New Players. The top two donors in 2011, the Bill & Melinda Gates Foundation with total giving of $448 million and the Walton Family Foundation at $160 million, rose rapidly to the number one and two positions.21 In 1998, the Walton Family Foundation was number 26 on the Foundation Center’s list of the 50 largest givers to K–12 education. The Gates Foundation, which was not founded until 2000, did not appear at all. Gates is now the largest foundation in the world with assets in 2014 of over $40 billion.22 The foundation has become the focal point for much of the current debate about how, and whether, philanthropists influence education reform.

“Higher-Leverage” Agendas. A key explanation offered by some researchers for the relative failure of foundation philanthropy to spur major education reform is that much of the giving has gone to support what schools already do, rather than to encourage systemic reform and significantly new ways of operating. Greene (2005) argued that too much philanthropy went for "lower-leverage" activities (e.g., providing training to educators, offering pedagogical or curricular innovations, giving schools additional equipment). Given the size of philanthropy relative to total school budgets, such gifts are like "buckets into the sea," too small to make a significant impact. He argued that “higher-leverage” activities, aimed at trying to redirect how future public expenditures are used, have more potential to effect real long-term systemic changes. Such activities include supporting research and advocacy efforts that inform education policy debates, creating new types of public schools or administrative structures through which public dollars will flow, and developing alternative professional associations and credentials.

Sarah Reckhow (2013) argues that there have been major changes in the role of education foundations over the past decade. Those changes include dramatically increased giving, more hands-on involvement, greater focus on policy advocacy, and more strategic investment targeting fewer recipients. In her study, Reckhow found that Gates and many other top education grant makers, such as the Broad Foundation and Carnegie Corporation, had dramatically increased their giving to K–12 education. More broadly, she finds that in 1998, the 50 largest K–12 donors distributed $328 million in grants for education; by 2010, it was nearly $1 billion. Reckhow also argues that in the last decade federal Department of Education initiatives and the agendas of independent foundations have largely converged.

A key example of this convergence is support for the Common Core standards. Private foundations, notably the Gates Foundation and the Hewlett Foundation, have heavily underwritten the development of the standards through support for the National Governors Association and the Council of Chief State School Officers, two highly influential state-based associations. The foundations also supported the development of tests—Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balance—designed to assess student performance relative to the standards. In addition, they supported advocacy efforts to promote widespread acceptance of the standards and professional development efforts to help ensure teachers’ readiness to teach to the standards. The Gates Foundation, based on its own reports (www.gatesfoundation.org), has spent over $175 million on projects related to the Common Core, most of it in 2013. Similarly, the Hewlett website (www.hewlett.org) reports that the foundation devoted over $14 million in Common Core support. In concert, the U.S. Department of Education promoted the Common Core by granting waivers for accountability provisions in NCLB and by identifying adoption of multi-state standards, such as the Common Core, as one of the key selection criteria for receiving a Race to the Top award.

Hands-on Involvement. Unlike the older foundations, whose founders and family members are often long gone, many of the newer and most visible foundations are actively managed by the business leaders who endowed them. The Broad Foundation (founded in 1999 with a focus on K–12 urban education) reflects “the
beliefs and core values” of founder Eli Broad in its efforts to improve education through better governance, management, labor relations, and competition (Hassett & Katzir, 2005). Laura and John Arnold (Arnold Foundation), Donald and Doris Fisher (Pisces Foundation), Bill and Melinda Gates (Gates Foundation), and John Doerr (NewSchools Venture Fund) are other examples of donors with active involvement in the decisions about the operations and beneficiaries of their charities.

Current philanthropy plays a larger role in policy advocacy than does traditional philanthropy. According to Reckhow (2013), founders tend to use the bully pulpit of their foundations for policy advocacy. As she points out, in 2010, "Gates delivered keynote speeches at the National Charter Schools Convergence, the American Federation of Teachers Convention, and the Council of Chief State School Officers Annual Policy Forum” (Reckhow, 2013, p. 29).

Grantmakers for Education, founded in 1995, today claims 280 member organizations including the Gates, Walton Family, and Arnold Foundations. One of the organization’s priorities is to “impact education policy at the local, state and national levels.” In 2012, Grantmakers for Education reported that 77 percent of its members, who fund policy work, stated they fund advocacy, up from 58 percent in 2008.

Strategic Philanthropy. These “new givers” are described as bringing “a flashier, more entrepreneurial, more aggressive approach to both giving money and insisting on results.” They are “more intrusive, ask much tougher questions, have a much more hands-on relationship. They bring a degree of impatience” (Colvin, 2005, pp. 29–30). They tend to work under theories of strategic and venture philanthropic giving—“the setting of clear goals, developing sound evidence-based strategies for achieving them, measuring progress along the way to achieving them, and determining whether you were actually successful in reaching the goals” (Brest, 2010). Unlike traditional philanthropy, which largely responds to schools’ and districts’ requests for support for ongoing projects, strategic philanthropy sets the objective itself. According to Brad Smith, head of the Foundation Center, in 2013, 60 percent of all American foundations—representing $16 billion in foundation giving—did not accept unsolicited proposals (Eisenberg, 2013). Education foundations such as the Gates, Broad, and Arnold foundations all follow the strategic grant-making philosophy, basing the foundation’s policy decisions on data, “and not just on instinct or fear” (Reagan, 2013).

Venture Philanthropy. Some foundations explicitly follow an approach called “venture philanthropy,” where private equity/venture capital models are applied in the nonprofit and charitable sectors. Venture philanthropy and venture capital investment are similar: both support start-ups, growth, and risk-taking ventures. The key difference is that venture philanthropists seek social, instead of monetary, returns. An example in education is the NewSchools Venture Fund (NSVF), which raises early-stage capital from a variety of institutional and individual donors and invests it in promising education entrepreneurs. One innovation that NSVF has invested in is New Schools for New Orleans—a nonprofit organization created solely to support the rebuilding of the public education system in New Orleans after hurricane Katrina.

Donor-Advised Funds. With assets in the billions and growing, donor-advised funds are proving a new way for smaller donors to provide additional resources to schools. These funds allow an individual organization to establish a small fund and direct a donation to eligible recipients. The donor will receive a tax deduction for placing the money in the fund, and the managing organization decides when to release the funds to the recipient. A variety of institutions including community foundations, federated giving programs, universities, and major financial institutions offer donor-advised funds. In education specifically, organizations such as Donors Choose allow public school teachers to describe a class or school project and request specific funding amounts from potential donors to support the project. Individuals and organizations can open an account with Donors Choose and fund any of the projects listed by the teacher. Founded in 2000, Donors Choose has raised over $236 million.

Collaboration and Convergence. Finally, foundations are increasingly collaborating with each other in order to focus their efforts and maximize the impact of their giving, consistent with theories of strategic and venture philanthropy. Grantmakers for Education found that since 2008 over 90 percent of grant funders have collaborated with other funders. This coordination can be seen in the funding for human capital developer, Teach for America. In 2011, Teach for America received over $73 million in grant funding, which included donations from over 179 different foundations, including the Broad, Walton, Gates, and Robertson foundations. At the local level, Colvin (2005) cites two superintendents (Alan Bersin in San Diego and Thomas Payzant in Boston) who worked to encourage foundations to avoid fragmentation by supporting a
district’s reform strategy rather than the foundations’ ad hoc projects and preferences. The Association of Baltimore Area Grantmakers’ (ABAG) mission is to maximize grant impact by networking and coordinating foundation funding for specific district projects. For example, in 2010, ABAG coordinated $11.9 million for Baltimore education programs, two-thirds of which went directly into the public school system.\textsuperscript{21}

\textbf{Challenges Facing Foundation Philanthropy}

Will new approaches to foundation philanthropy be more successful than old ones in spurring significant K–12 improvement? It is too early to tell, but what is clear is that foundations are breaking away from the historic methods of education support and operating under a corporate-type model of strategic reform. Frumkin (2001) has asked whether the new venture philanthropy will really turn out to be substantially different than its forebears or whether the apparent changes in operating style are more a matter of semantics than reality (with, e.g., “investments” equating to “grants,” “investors” to “donors,” “social returns” to “impact,” “performance measurement” to “evaluation,” “due diligence” to “grant review process,” and so on).

Hess (2005) and Reckhow (2013) identify some similar and ongoing challenges raised by philanthropic involvement in efforts to influence the course of America’s schools.

- \textit{“Strategic” giving}: How donors can advance strategic visions of reform without stifling experimentation and diversity of ideas and coalitions involved in reforms.
- \textit{Philanthropy and accountability}: How to obtain good and transparent evaluations in the face of such political realities as donor reluctance to incur criticism and researchers whose present or future prospects may be tied to donor support.
- \textit{Pipelines versus programs}: How to increase donor appreciation of the value of pipelines that seek to alter the composition of the education workforce by attracting and/or keeping new lines of talent, as contrasted to programs that focus on a limited population of children and educators. Pipelines (examples include Teach for America, Troops to Teachers, and the National Board for Professional Teaching Standards) have less immediate and direct impacts on individual schools than programs such as reading programs or computer labs, but they offer longer-term levers for systemic human capital change.
- \textit{Balancing performance and patience}: How to reconcile the new focus on disciplined management strategies and measurable results in the near term with the reality that worthwhile efforts may not demonstrate near-term effects but can more broadly contribute to reshaping education.

Both Hess (2005) and Reckhow (2013) also make the point that, as foundations engage in higher-leverage giving intended to influence public policy, they should expect to be held to higher standards of public accountability. Just what such standards might be, and how they might be enforced, however, are unclear.

\textbf{Corporate Supporters}

America’s business community has had an off-again, on-again relationship with the nation’s public schools. Beginning with the West Indies Company, which started schools in New Amsterdam (later New York City) in the 1640s to attract settlers (Lenkowsky, 2005), businesses were important in founding schools for America’s children. After publicly provided education became widespread, business involvement shifted from the creation to the support of schools, both financially and through such activities as service on school boards. While many business leaders pulled back from K–12 education during the politically tumultuous 1960s and 1970s, they began to reengage again in the 1980s in response to global economic changes that focused attention on the quality of the nation’s workforce. Reinvigorated business involvement with the schools was accompanied by a shift in corporate contribution programs toward a new emphasis on precollegiate education (Timpane & McNeill, 1991).

As with other kinds of private givers, it is difficult to portray succinctly the myriad of educational activities...
of the diverse business community. There is a continuum of involvement, and individual businesses and business people work with schools for reasons both altruistic and self-serving. Some key points related to corporate support can most effectively be conveyed by a brief discussion of two types of activities: philanthropic support and commercial partnerships.

**Corporate Philanthropic Support of K–12 Education**

Businesses small and large contribute to schools in a variety of ways, both financial and nonfinancial. Employees donate time to work in classrooms and mentor students or teachers. Their employers give equipment and money to individual schools and districts, sometimes directly and sometimes through LEFs and other community groups. Tim- pane and McNeill (1991) describe four “patterns” of business involvement:

- **Helping-hand relationships:** These encourage businesses to supplement or enhance existing school programs through donations of tangible goods and services such as equipment, mini-grants, tutors, speakers, and special materials.
- **Programmatic initiatives:** More intensive, more complex partnerships between businesses and schools aimed at improving one particular program or school through, for example, career academies and mentoring programs.
- **Collaborative and compacts:** Joint efforts involving several businesses and one or more school districts, such as the LEFs mentioned earlier.
- **Policy change:** Through their involvement with such organizations as the national Business Roundtable and its state affiliates, the Committee for Economic Development and various state and city business associations, business leaders have engaged in public policy debates over higher standards, accountability for outcomes, and restructured schools.

Corporate financial giving to K–12 education is done both directly and through corporate foundations, which are private foundations that receive their assets from a company rather than an individual or family. These foundations may have an endowment, but the bulk of their giving generally comes in the form of annual gifts from the company. Companies may give both directly and through their foundations. The latter are subject to IRS regulation; the former have no public disclosure requirements, though the Foundation Center tries to include them where it can. According to the Foundation Center (2014a), 2,700 corporate foundations in 2012 gave away $5.2 billion, representing 11 percent of all private philanthropic giving. Of all corporate foundation giving, 20 percent was directed to education at all levels, not just K–12.

Corporate foundations face many of the same criticisms and challenges that independent foundations do in terms of the impact their activities have on creating a more effective K–12 education system. Rotherham (2005) suggests that corporate foundations are more inclined than their independent brethren to invest in lower-leverage rather than higher-leverage activities. Frumkin (2005) finds corporate philanthropists more likely to make gifts in a relatively detached way, in order to avoid exposure to negative publicity if their charitable programs backfire or fail. Corporate donors are also more apt to be detached because their foundations and giving arms often have insufficient staff to permit deeper involvement.

**Commercial Partnerships**

One type of school-oriented business activity that has grown in size and visibility over the last 25 years involves explicitly commercial relationships. The change has been so noticeable and controversial that Congress asked the General Accounting Office (now the Government Accountability Office, GAO) to undertake a study looking at the nature and extent of commercial activities in schools and the laws, regulations, and policies that govern them (GAO, 2000).

GAO found that school-business commercial partnerships had been growing because of limited school budgets, an increased demand for educational services, and the growing purchasing power of American youth. GAO (2000, p. 5) described the benefits perceived by schools and businesses from the increase in commercial arrangements:
In general, schools want cash, equipment, or other assistance in providing services and technology during a period when revenues from traditional tax sources are, for many school districts, essentially flat. Businesses want to increase their sales, generate product loyalty, and develop climates favorable to their products, although some businesses are involved with schools primarily to help local communities.

GAO classified the commercial activities it found in a sample of schools as product sales, direct advertising, indirect advertising (e.g., via corporate-sponsored educational materials, teacher training, contests or incentives, and grants and gifts), and in-school market research. Product sales were the most common and lucrative type of school-oriented commercial activity, especially the sale of soft drinks.

Advertising appeared especially on soft drink machines and high school score boards, though some was delivered to classrooms via media (e.g., Channel One). Indirect advertising was “limited and subtle,” while in-school market research was nonexistent in the sampled schools (though one principal had been approached about doing research). GAO found that high schools had more commercial activities than middle or elementary schools, that state laws and policies were spotty and varied, and that most decisions about whether to engage in commercial activities were left up to local officials. The revenues generated by commercial activities were tiny percentages of district revenues, but their intensity varied from place to place.

Commercial activities such as soft drink contracts (which guarantee schools some amount of per-pupil revenues or a percentage of sales) and electronic advertising in classrooms have been especially controversial. Due to concerns about rising levels of obesity in children, soft drink contracts have been under attack (Strom, 2012). They have raised the ire of groups opposed to the overcommercialization of society and concern about the effects of advertising on young children. What such developments will mean for school revenues depends on the popularity of the substitute products that the companies will provide.

Concluding Thoughts

The universe of nongovernmental organizations providing financial resources to public schools is diverse and fragmentary. While the total amounts of revenue involved are very small compared to government funding of elementary and secondary education, in “niches” private funders can have an important impact both for good, such as when they spur important improvements in educational offerings or in the operation of the system, or for ill, as when they encourage unhealthy habits in children. As private donors engage in more systematic efforts to spur major reforms in America’s educational enterprise, their actions take on aspects of public as well as private interest. Thus, enhanced efforts to accumulate better data and understanding of the role of nongovernmental organizations in financing public schools are of ongoing importance.

Acknowledgments

The original version of this chapter was authored by Janet Hansen. This revised version incorporates contributions by Michelle Hall, Dominic Brewer, and Jane Hannaway. The research assistance of Jeffrey Nelson is gratefully acknowledged.

Notes

2. For example, districts may receive tuition payments from various sources and rental fees from patrons who use school facilities.
3. See, for example, Zimmer, Krop, and Brewer, 2003; Addonizio, 2000; Brunner and Sonstelie, 1997; Brunner and Imazeki, 2005.
5. The F-33 form used by the Census Bureau to collect finance data on school systems is online at
The accounting codes used on the F-33 form are from National Center for Education Statistics (NCES), 2009. NCES also collects information on public school revenues by source on the National Public Education Finance Survey (ED Form 2447), but this survey is based on statewide rather than district-level data; and the “other revenue from local sources” category in which contributions and donations from private sources are to be reported includes an even larger number of other revenue types than does the F-33.

For example, Lampkin and Stern (2003, p. 17) report that fewer than 40 percent of the members in the California Consortium of Education Foundations are represented in a database of organizations that file Form 990s.
References


Introduction

All 50 state constitutions mandate a statewide system of public education, charging state governments with providing a "thorough and efficient system of free public schools" (or similar language). State legislatures fulfill this obligation by establishing compulsory schooling laws, curriculum standards, and institutions governing the formation and operation of school districts. Despite their constitutional mandate, states historically played only a minor role in the financing of public schools, leaving the funding and day-to-day management of schools in the hands of local governments.

This long-standing devolution of fiscal responsibility to local school districts all but dissolved over the past century. State aid in the form of flat grants and "minimum foundation" programs became common practice in the first half of the 20th century; later, legal challenges to state school finance systems led to far-reaching reforms that further expanded the fiscal responsibilities of the states. As shown in Section III of this handbook, these challenges originated as "equity" suits that sought to break the link between local property wealth and school resources. A later wave of litigation instead sought "adequacy" in school expenditure, as defined by the level of spending necessary to reach some performance standard. Though the legal grounds for these challenges differed, the result was often the same: an increased state role in school funding. On this matter, states frequently had few alternatives, as most major reforms to state school finance systems were legislative responses to a court order. In several cases, though, legislatures enacted significant reform without the prompting of the courts. Beyond direct aid to elementary and secondary schools, states also indirectly affect education spending through policies that shape local taxes and spending, such as the setting of minimum tax rates or mandates in service provision.

In this chapter, we examine the evolving role of the states in financing primary and secondary education in the United States. We begin in the next section by documenting the increased importance of states as a source of revenue for schools, show how reliance on state funding varies across states and within states over time, and explore some of the key explanations for these trends. Following that, we examine the extent to which finance reform has increased the states’ role in education funding, and how in turn these reforms have affected the level and distribution of per-pupil expenditure between and within states. Finally, prior to a brief concluding section, we review how an increased state role in education finance has affected various nonfiscal aspects of the public education system, most importantly student outcomes.

Education Finance and the Role of the States

U.S. education’s early roots in the English tradition of locally provided private or religious schools influenced the division of fiscal responsibility for education for nearly two centuries. Not until the mid-19th century “common school” movement did states formally recognize public education in their constitutions, creating statewide systems of public schools to be funded almost entirely from local revenue sources (Odden & Picus, 2004).
In the 1920s, local taxes accounted for the vast majority of funds for public education (83.2 percent, see Figure 21.1) with the remainder coming from states (16.5 percent), mostly through flat per-student grants. During these years, federal contributions to public education were negligible at less than 1 percent of all school revenues. The 1930s saw a doubling in the state share of school expenditure, as state governments sought to assist localities whose property tax base had eroded during the Depression with “minimum foundation” programs (Benson & O’Halloran, 1987; see also Picus, Goertz, & Odden, this volume). The state share of revenues continued to rise in the 1940s as localities coped with the effects of the baby boom on primary school enrollment, but stalled at just under 40 percent of public school expenditures for the next 25 years. Meanwhile, the 1965 passage of the Elementary and Secondary Education Act caused the federal contribution to school spending to jump from 4.4 percent in 1960 to 8 percent in 1970, and nearly 10 percent in 1980. For the next quarter century, however, the federal share remained below 8 percent.

As shown in Figure 21.1, the state share of public K–12 expenditure surged in the 1970s, rising from 39.9 percent in 1969–70 to 46.8 percent by the end of the decade. Perhaps surprisingly, the aggregate state share has remained between 45 and 49 percent since 1978—despite reforms of the 1980s and 90s that vastly increased state funding in some states. Still, in only four of 33 years since 1978 have local school districts contributed more to education revenues than have state governments. They have done so only once since 1994, in 2010, when the Great Recession led to unprecedented cuts in state aid. The federal share began to rise in the 2000s, following No Child Left Behind, and surged to 12–13 percent in 2010 and 2011 as the federal government channeled stimulus funds to states in response to the fiscal crisis (see Gordon, this volume, for more on the evolving role of the federal government in education finance). Between 2007–08 and 2009–10, federal aid per pupil rose 53 percent in real terms.

The national trend in state funding masks considerable variation across states in aid to local school districts. In 1970, for example, the fraction of school revenues from state sources varied across states from a low of 10–18 percent in New Hampshire, South Dakota, and Nebraska to a high of 60–70 percent in New Mexico, Alabama, and Delaware. By 2010, the variability across states in the state share was much smaller, ranging from lows of 29–32 percent to a high of 65–83 percent. Almost every state increased its share of public school revenues over this period, with Florida, Louisiana, and Pennsylvania being notable exceptions. Generally speaking, increases in the state share appear to reflect increases in state spending on education rather than declines in local revenues.
Explanations for this dramatic shift in fiscal responsibility for public schools from local to state can be found in part in a series of legislative and court-mandated reforms to state finance systems that originated in California’s *Serrano I* ruling (1971) and continue to this day. These reforms arose initially in response to inequities generated by reliance on local property taxes for school funding, but in recent years have been driven by concerns over the adequacy of funding for public education, in particular the funding of education for disadvantaged students.4

Pressure to reform school funding originated in both state legislatures and in the courts. At last count, litigants had challenged the constitutionality of state school finance systems in all but five states. As Table 21.1 indicates, state supreme courts have ruled on the legality of school funding systems in 40 states since 1970, with 19 declaring their system to be unconstitutional and 30 upholding the existing state program (nine states have had both types of ruling; see Appendix Table 21.A.1 for a detailed inventory). In most cases, a decision by a high court to overturn a state education financing system has been accompanied by a direct order to make fundamental changes to school funding formulas, an order to which the legislature has responded in full (e.g., Kentucky, Washington), in incremental steps (New Jersey), or barely at all (Ohio). Legislatures have also initiated their own far-reaching reforms to school finance systems, in the wake of unsuccessful litigation (e.g., Georgia and Idaho), under the threat of litigation (e.g., Missouri and Oklahoma; see Minorini & Sugarman, 1999), or in response to political pressure (e.g., Michigan). Appendix Table 21.A.1 provides our categorization of court rulings as “equity” or “adequacy” based, although some legal scholars have argued that this distinction is not always clear (Briffault, 2007).

While much of the literature has focused on the effects of these finance reforms, a smaller body of work has examined the circumstances under which reform arises. Baicker and Gordon (2006), Figlio, Husted, and Kenny (2004), and Card and Payne (2002) all demonstrate that high-court rulings overturning school funding systems are difficult to predict, even after researchers account for variation in the constitutional language upon which such rulings are purportedly

---

**Table 21.1 State Supreme Court Rulings on School Finance Constitutionality**

<table>
<thead>
<tr>
<th>State</th>
<th>Case</th>
<th>Year</th>
<th>State</th>
<th>Case</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arkansas</td>
<td>Tyler</td>
<td>1962</td>
<td>Arizona</td>
<td>Hufsted</td>
<td>1972</td>
</tr>
<tr>
<td>Louisiana</td>
<td>1803</td>
<td>1862</td>
<td>Colorado</td>
<td>Janis</td>
<td>1882</td>
</tr>
<tr>
<td>Connecticut</td>
<td>Hilty</td>
<td>1919</td>
<td>Illinois</td>
<td>1988</td>
<td></td>
</tr>
<tr>
<td>Kansas</td>
<td>Sheldon</td>
<td>1950</td>
<td>Kansas</td>
<td>1971</td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td>Iran</td>
<td>1899</td>
<td>Maine</td>
<td>1951</td>
<td></td>
</tr>
<tr>
<td>Louisiana</td>
<td>Muth</td>
<td>1861</td>
<td>Maryland</td>
<td>1970</td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>McLaugh</td>
<td>1919</td>
<td>Maryland</td>
<td>1970</td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>Holland</td>
<td>1919</td>
<td>Maryland</td>
<td>1970</td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td>Bloomfield</td>
<td>1969</td>
<td>Missouri</td>
<td>1973</td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td>Minnesota</td>
<td>1973</td>
<td>Missouri</td>
<td>1973</td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td>Minnesota I</td>
<td>1989</td>
<td>Missouri</td>
<td>1973</td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td>Minnesota II</td>
<td>1990</td>
<td>Missouri</td>
<td>1973</td>
<td></td>
</tr>
<tr>
<td>New Jersey</td>
<td>1969</td>
<td></td>
<td>New Jersey</td>
<td>1969</td>
<td></td>
</tr>
<tr>
<td>New Jersey</td>
<td>1972</td>
<td></td>
<td>New Jersey</td>
<td>1972</td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>ACE</td>
<td>1975</td>
<td>New York</td>
<td>1975</td>
<td></td>
</tr>
<tr>
<td>North Carolina</td>
<td>Lineberger</td>
<td>1979</td>
<td>North Carolina</td>
<td>1979</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>Sheldon</td>
<td>1979</td>
<td>Ohio</td>
<td>1979</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>Sheldon</td>
<td>1979</td>
<td>Ohio</td>
<td>1979</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>Sheldon</td>
<td>1979</td>
<td>Ohio</td>
<td>1979</td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>1977</td>
<td></td>
<td>South Carolina</td>
<td>1977</td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>Small School I</td>
<td>1979</td>
<td>Tennessee</td>
<td>1979</td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>Small School II</td>
<td>1979</td>
<td>Tennessee</td>
<td>1979</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>Edmondson</td>
<td>1988</td>
<td>Texas</td>
<td>1988</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>Edmondson</td>
<td>1988</td>
<td>Texas</td>
<td>1988</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>Edmondson</td>
<td>1988</td>
<td>Texas</td>
<td>1988</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>Edmondson</td>
<td>1988</td>
<td>Texas</td>
<td>1988</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>Edmondson</td>
<td>1988</td>
<td>Texas</td>
<td>1988</td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>Edmondson</td>
<td>1988</td>
<td>Texas</td>
<td>1988</td>
<td></td>
</tr>
</tbody>
</table>

*Source:* See Appendix Table 21.A.1
Consequently, empirical research investigating their effects typically treats court-mandated reform as an exogenous event. Reforms initiated by the legislature cannot, of course, be viewed in this way and perhaps for this reason have been less frequently studied as a group.  

Several authors have examined the growth in state aid to local school districts more comprehensively, relating redistributive state aid to state demographics and political measures, in addition to exogenous court orders. De Bartolome (1997), for example, attributes much of the growth in state aid between 1970 and 1990 to rising income inequality. His logic is that a rise in income inequality raises mean income relative to that of the median, or decisive, voter in the political process and encourages a greater redistribution from high-income to low-income school districts through state aid. Recognizing that the same level of state aid can be distributed in more or less targeted ways, Figlio, Husted, and Kenny (2004) instead relate within-state revenue inequality to various state characteristics. Holding constant the court rulings on school finance constitutionality, these authors find fewer spending disparities (i.e., greater redistributive state aid) in states with stronger constitutional language toward equity and in states under Democratic control, and greater disparities in states with more heterogeneous populations.

Education Finance Reform and the Level and Distribution of Spending

A natural question to ask is whether education finance reforms have generated changes in the level and interdistrict distribution of school resources within affected states. There are good reasons why one might not observe expected changes, even as courts have ordered and states have enacted policies specifically targeting these objectives. State aid formulas vary widely (Loeb, 2001; Hoxby, 2001; Fernandez & Rogerson, 2003; Yinger, 2004) and, in the case of judicially mandated reforms, may be crafted with an eye toward satisfying court requirements rather than institutionalizing incentives that yield desired outcomes. Reforms deemed to meet constitutional requirements may in fact embody incentives that yield unintended consequences. It thus remains an empirical question whether these reforms have, in practice, affected the level and distribution of school funding.

In this section, we assess how finance reform has altered the role of states in funding public education, and in turn how these reforms have affected the level and distribution of school resources. We focus primarily on court-ordered reform, but briefly address the literature on the effects of reforms originating in the legislature. In conducting our analysis, we have chosen not to adjust reported school expenditure for differences in student populations or other district-specific cost factors, because there is little consensus in the literature about how to adjust for cost differences (see Duncombe & Yinger, this volume), the weights on various cost factors may have changed over time, and the data required to carry out comprehensive cost adjustments over a four-decade period are not readily available.

Trends in Per-Pupil Spending Inequality

In Table 21.2, we use several datasets of school districts spanning the 1972–2011 period to compute four measures of inequality in per-pupil current school expenditures, using school districts as the unit of analysis, and weighting by district enrollment. For years 1990 and before we rely on a panel of unified school districts originally compiled by Harris, Evans, and Schwab (2001) that includes revenue and expenditure data from the Census of Governments. For years 1992 and following we use annual F-33 Census survey of school districts data. The first measure of interdistrict spending inequality—the ratio of the 95th percentile of per-pupil expenditure and the 5th percentile—captures inequality by comparing the extremes of the spending distribution, while the other three (the Gini coefficient, Theil index, and coefficient of variation) provide measures of inequality that take into account the entire distribution of spending. The Theil index is particularly useful in that it can be decomposed into disparities within and between states. All of these measures increase as the level of inequality in expenditure increases.

Table 21.2 Inequality in Per-Pupil Current Expenditures, United States, 1972 to 2011
All measures of inequality in Table 21.2 follow a similar pattern over time. Inequality in per-student funding fell sharply between 1972 and 1982 and rose steadily during the 1980s. The 1990s saw an almost monotonic decline in inequality—as much as 18 to 22 percent, depending on the measure (the Theil index suggests an even greater decline of 34 percent). Recent years, however, have witnessed a sharp uptick in inequality—the Gini coefficient, for example, rose from 12.4 in 1999–2000 to 16.3 in 2010–11, a 32 percent increase. This level of the Gini is as high, or higher, than at any other point since 1971–72.

Our decomposition of the Theil index into between and within state components sheds additional light on school spending inequality in the United States. First, in most years, 60–70 percent of the variation in spending can be attributed to between-state differences in per-pupil expenditure. Second, declines in between-state inequality accounted for more than half of the drop in overall school spending inequality between 1972 and 2004. This may be partially due to income convergence across states, as variation in state per capita income follows a similar time series pattern to that of between-state inequality in school funding—sharp declines in the 1970s, followed by a sizable increase in the 1980s and a moderate decline in the 1990s (Bernat, 2001; see also Hoxby, 1998). Finally, both within-and between-state inequalities have risen in recent years, with between-state inequalities accounting for the lion’s share (about 90 percent) of the 2000–11 rise in the Theil index. Reschovsky (2004) and Kalambokidis and Reschovsky (2005) attribute the 2000–04 rise in inequality and parallel decline in the state share of school revenues to a post-recession fiscal crisis that led to sharp cuts to state education aid. The larger, more recent surge in between-state school spending inequality reflects the disproportional impact of the Great Recession on state budgets (Dadyan, 2012).
Time series trends in aggregate spending disparities mask considerable differences within states. Figure 21.2 plots each state’s Gini coefficient of per-pupil spending inequality in 2006–07 against its Gini in 1971–72, with points weighted by their relative 1972–2007 growth in real per-pupil current expenditures. (We use 2006–07 as an end point to avoid any influence of the Great Recession.) Measured by the Gini, most states experienced a decline in interdistrict disparities between 1972 and 2007, evident in Figure 21.2 by a data point below the 45-degree line. Some states saw sizable reductions in inequality—North Carolina, California, Delaware, and Washington, for example, saw a 40–50 percent decline in within-state spending inequality. Others—such as Arizona, Massachusetts, and Idaho—became more unequal. A comparison of state Gini coefficients in 1990 and 2004—an era notable for courts’ emphasis on resource adequacy rather than equity—also shows evidence of equalization in spending, though reductions in inequality were less dramatic (states such as Georgia, Oregon, Maryland, and Texas, however, became substantially more equal).

A recurring question in the school finance literature is whether or not finance reforms achieve their goals by “leveling up” or “leveling down” spending on education (Manwaring & Sheffrin, 1997; Hoxby 2001). "Leveling up" may be loosely defined as a reduction in spending disparities accomplished by a rise in spending in historically low-spending school districts, a condition under which average spending increases. "Leveling down," in contrast, is a reduction in spending disparities accomplished by reducing spending at the top of the distribution, either absolutely or relative to what it would have been in the absence of reform. In Figure 21.2, a strong leveling-down effect would be apparent if expenditure growth were lower (the point size were smaller) the further the state appears below the 45-degree line. In fact, this does not appear to be the case, although there are some possible exceptions (California, North Carolina, Washington, and Florida, for example). Of course, these correlations are only suggestive. We turn next to a more systematic analysis of the effects of court-ordered finance reform on school spending.

The Fiscal Effects of Education Finance Reform—Evidence From Specific States
As the first and arguably the most aggressive state effort to equalize education spending across districts, California’s response to *Serrano v. Priest* (1976) has received considerable attention. The post-*Serrano* reduction in spending disparities in that state was accompanied by a significant reduction in real spending growth, with mean expenditures per student dropping from 11th to 30th in the nation (Silva & Sonstelie, 1995). This collapse in expenditure growth was attributable in part to Proposition 13, the 1978 property tax limitation that followed *Serrano* (see Downes & Figlio, this volume), and Fischel (1989, 1996) convincingly argues that voter support for Proposition 13 was a direct reaction to the *Serrano* reforms. Ultimately, Silva and Sonstelie attribute more than half of California’s “leveling down” to the effects of *Serrano*, and the remainder to enrollment growth caused by immigration. Coen-Pirani (2011) estimates that spending per student in California would have been 24 percent higher in 2000 if immigration had remained at its 1970 level.

While California has become the most visible example of the fiscal effects of finance equalization, most scholars of school finance view it as an exception to the rule, or at least an extreme case. School finance reform in many other states appears to have been accomplished through a leveling up of overall spending. For example, Roy (2011) and Cullen and Loeb (2004) document both a sharp reduction in spending inequality in Michigan after its 1995 reform and a growth in mean per-pupil expenditure that surpassed the national average. Clark (2003) and Flanagan and Murray (2004) found considerable equalization in current expenditure after Kentucky’s 1990 reform responding to *Rose v. Council for Better Education, Inc.*, yet Kentucky ranked third in the nation in per-pupil expenditure growth between 1990 and 2004.

The effects of finance reform on expenditure occur through a variety of channels (Manwaring & Sheffrin, 1997; Fernandez & Rogerson, 2003; Downes & Shah, 2006). For example, increased centralization of school finance may alter the public choice mechanism relevant for the determination of school spending. Under local finance, households sort into communities based on their demand for school quality, which is determined in part by income. A fully centralized system effectively creates a single jurisdiction. In a simple collective choice model in which spending is determined by the median voter, one would predict lower school expenditure (i.e., a “leveling down”) under a centralized system than under a decentralized system if median state income lies below the mean. Hoxby (2001) shows that when local school districts retain some discretion over expenditure, a reformed state aid formula can reduce incentives to spend at the top of the expenditure distribution more than it raises the incentive to spend at the bottom, again resulting in a leveling down. Aid formulas may contain additional features that result in a mechanical leveling down of expenditure. Hoxby and Kuziemko (2004) argue that the “recapture” feature of the “Robin Hood” formula in Texas destroyed property wealth in rich districts, upon which the system depended. Finally, leveling down may arise as a long-run “general equilibrium” response to finance reforms, such as tax revolts, increased private school enrollment, and political fractionalization undermine support for public education (Nechyba, 2004).

Given the diversity of finance systems, political and constitutional histories, demographics, and preferences for public education across states, it is reasonable to ask how finance reforms on average have affected within-state inequities and the overall level of spending on public schools. We turn next to this question.

### The Fiscal Effects of Education Finance Reform— Evidence Across States and Over Time

Estimates of the average impact across states of school finance reform typically involve some variation on the following empirical model (see Murray, Evans, & Schwab, 1998; Evans, Murray, & Schwab, 1997 and 1999; Corcoran et al., 2004; Berry, 2007; Springer, Liu, & Guthrie, 2009; and Sims, 2011a):

\[
y_{it} = X_{it} \beta + D_{it} \alpha + \mu_i + \eta_t + \epsilon_{it} (1)
\]

where

- \(y_{it}\) is some measure of per-pupil spending or spending inequality for state \(i\) in year \(t\)
- \(X_{it}\) is a vector of state demographic characteristics
- \(D_{it}\) is an indicator of the status of finance reform in state \(i\) in year \(t\)
- \(\mu_i\) and \(\eta_t\) are state and year fixed effects, respectively, and
- \(\epsilon_{it}\) is an idiosyncratic error

315
State fixed effects allow for the possibility that reform states differ in systematic ways from nonreform states; for example, in the form of differences in existing spending equality or school governance. Year effects control for secular trends in the dependent variable common to all states, and the demographic controls (such as for age, race, or income) account for correlates with school expenditure that may also influence reform legislation. Taken together, in the case where \( y_t \) is spending inequality and \( D_t \) is an indicator that equals one in all subsequent years after reform, \( \alpha \) can be interpreted as the average impact of finance reform on within-state spending inequality. \(^{21}\)

Following Murray, Evans, and Schwab (1998) and Corcoran et al. (2004), we use state-level data from the Digest of Education Statistics (2005) together with a balanced panel of more than 10,300 school districts to further extend this analysis through 2002. \(^{22}\) We estimate model (1) for four separate measures of within-state spending inequality and for per-pupil spending at various points in the state expenditure distribution. State and year fixed effects are included along with an indicator variable showing whether or not the school finance system in that state had been declared unconstitutional in the current or an antecedent year. Only high-court rulings on constitutionality are considered (those in Table 21.1), a slightly more stringent inclusion criterion than that followed in past literature.

Table 21.3 Impact of Court-Mandated Education Finance Reform on the Level and Distribution of School Funding, 1971–72 to 2001–02

Table 21.3 summarizes the estimated effects of court-mandated school finance reform on the level and distribution of spending within states. In this table, each row summarizes the results of a separate regression, where the dependent variables used are listed in the left-hand column. Column (1) presents the estimated coefficient on our reform indicator \( (D) \) from each regression, while column (2) provides results from an alternative specification of \( D \) (described below). In the following discussion, the marginal effects reported in Table 21.3 have been converted to percentages using the baseline (1972) mean values of the relevant dependent variable.

With 1972 as a baseline, we find that within-state inequality fell as much as 15–19 percent between 1972 and 2002 in states with court-mandated finance reform, relative to those without such orders (Panel A, column
The decrease is considerably higher, at 36.7 percent, for the Theil index. All of these estimates are statistically precise, and are comparable to those in Murray, Evans, and Schwab (1998) and Corcoran et al. (2004), who found a 19–34 percent decline in inequality in reform states, relative to nonreform states. Not surprisingly, the average effect on inequality over the full 1972–2002 period examined here is somewhat smaller than that found in earlier work covering a shorter time horizon; these smaller estimates are likely attributable to the preponderance of adequacy (versus equity) rulings in the new years of data. Berry (2007) and Springer, Liu, and Guthrie (2009) also examined the effects of court-mandated reform on within-state equity over this period. Although their criteria for defining court rulings and empirical specifications differ somewhat from ours, their findings are not notably different from those in Table 21.3.

Table 21.3 also summarizes estimates of the average effect of court-ordered finance reform on the level of spending, on average and at various points in the spending distribution (Panel B column (1)). During this period, the level of spending per pupil rose by a statistically significant 9.2 percent more in states under judicial orders than in states without a court order. Further, as found in earlier work, we see that spending in reform states grew more at the bottom of the distribution than at the top. After a high-court order, spending per student at the 5th percentile in reform states grew 12.5 percent more than in comparably low-spending districts in nonreform states. Spending grew 10 percent more at the median, and 5.5 percent more at the 95th percentile relative to states without reform orders, with the latter result statistically indistinguishable from zero. Taking all of the estimates in column (1) of Panels A and B together, we find little evidence of “leveling down” after court-mandated reform—at least in the aggregate—consistent with Murray, Evans, and Schwab (1998), Corcoran et al. (2004), and Berry (2007).

A natural extension of model (1) tests for differences in the impact of court rulings made on the basis of equity and those made on adequacy grounds. Our comparison of results in Table 21.3 to prior literature suggests that court-ordered reforms based on adequacy may have had weaker effects on spending disparities than those related to equity. One might also expect adequacy rulings to have more of a “leveling-up” effect on spending than equity orders. In column (2) we reestimate model (1) with an added indicator variable denoting that a court ruling was based on adequacy. The estimated coefficient on this variable may be interpreted as the differential effect of adequacy versus equity rulings.

Not surprisingly, estimates of the effect of court-ordered reform on spending disparities are smaller for rulings based on adequacy. For every measure, we find that adequacy-based rulings reduced spending inequality by roughly half that of equity rulings. These point estimates are similar to those in Berry (2007), and Springer, Liu, and Guthrie (2009), although the former is unable to reject the hypothesis of equivalent effects. When considering the level of expenditure, adequacy rulings appear to increase spending to a greater extent than equity rulings at all points in the distribution, though these differences are not always precisely estimated. These rulings have had much less of an impact on within-state inequality, however.

Sims (2011a) takes this further by asking whether adequacy judgments result in a better targeting of resources to districts with the greatest needs. He notes that adequacy reforms can increase spending more in high-educational-need school districts—while leaving traditional measures of equity unchanged—if high-needs students are not necessarily located in historically low-spending districts. He finds that adequacy-based court rulings significantly weaken the oft-observed negative relationship between poverty and school spending.

We also examine the local and state responses to court-ordered finance reform by looking separately at the effects of court-ordered reform on revenues by source. Consistent with earlier work, we find that states under court-ordered reform on average increased revenues per pupil by $701 (or 16.4 percent) more than nonreform states (Table 21.3, Panel C column (1)). Further, this increase appears to have been financed entirely by state funding, which rose on average $786 more per pupil in reform states. If anything, localities reduced their own funding in response to court-mandated reform, though we cannot rule out the hypothesis of no reduction in local funds. Taken together, the net effect of court-mandated reform appears to have been a 6.7 percentage point increase in the state share of education funding, relative to nonreform states.

In a separate analysis (not shown), we apply model (1) to annual data from the 1989–2004 “adequacy” era. Over this period, we find the average impact of court-mandated reform on within-state spending disparities to be statistically indistinguishable from zero. Surprisingly, court orders based on equity had a larger effect on spending in reform states than did adequacy-based rulings. The level effects of adequacy rulings are all estimated to be positive (if imprecise), and fairly uniform across the spending distribution.
The estimates provided here of the equalizing or level effects of school finance reform have relied exclusively on court rulings as an exogenous shock to the state’s system of school funding. A related literature has sought to move beyond “black box” estimates of the effects of finance reform by exploring the channels by which fiscal reform impacts spending. Manwaring and Sheffrin (1997) and Downes and Shah (2006) each interact reform indicators with key state-level determinants of school expenditure, recognizing that reform may fundamentally alter these underlying relationships. For example, as discussed above, Silva and Sonstelie (1995) pointed out that centralization may change the relationship between income and school spending in a state, as expenditure adjusts to reflect the demands of the median versus mean income household. Both Manwaring and Sheffrin (1997) and Downes and Shah (2006) find some evidence to support the hypothesis that a state’s response to fiscal reforms will depend to some degree on its own population characteristics.

Evans, Murray, and Schwab (1997), Manwaring and Sheffrin (1997), and Downes and Shah (2006) further contrast the effects of reforms that were a response to court mandates with those originating in legislatures. Court-mandated reform is reputedly more stringent and more restrictive of local discretion, and thus may be more likely to “level down” education spending. Indeed, each of these papers found evidence of a stronger impact of court-mandated reform on reducing spending disparities than legislative reform. Of course, it is more difficult to treat legislative reforms as exogenous events to which spending changes can be attributed, so such findings should be interpreted with caution.

Finally, an alternative approach taken by Hoxby (2001) avoids the legislative versus court-mandated reform typology and looks to state aid formulas themselves to predict whether a finance program will level up or level down school expenditure. Because aid formulas in effect determine each district’s “tax price” of per-pupil school expenditure, finance reforms can significantly alter a district’s incentives to spend on education. How a reform may be expected to affect the level and distribution of spending in a state will depend how—on net—districts respond to changes in their tax price. Hoxby’s empirical estimates show that states with the most dramatic equalization (such as California, New Mexico, South Dakota, and Utah) had much lower per-student expenditure than would be expected under a finance scheme with less distortionary tax prices, or under a purely local system of finance.

**Effects of Education Finance Reform on Other Public Expenditures**

Model (1) has also been used to estimate the effects of finance reform on other public expenditures at the state and local level. As shown above, court-ordered reform on equity and adequacy grounds has increased school spending, particularly at the low end of the distribution. Further, the increase in revenues necessary to support this growth has occurred largely through increased state funding with only a minimal decline in local funds. As state governments have increased their aggregate contribution to public education, one could ask whether this increase has occurred at the expense of other public programs.

Murray, Evans, and Schwab (1998) look at the impact of increased state support of public education on welfare, health, corrections, transportation, and higher education, and find no evidence of a reduction in per-capita expenditure in any of these categories. In contrast, Baicker and Gordon (2006) found over a longer time period that states reduced aid to localities for other public programs in response to court-ordered reform, and to a greater extent in counties with lower median income. They estimate that each additional dollar of state aid for education crowded out 22 cents in aid for public programs such as welfare, hospitals, and transportation. They further show that local governments respond to higher state taxes—necessary to support increased state aid to schools—by reducing their own taxes and expenditure (see also Berry, 2007).

**Nonfiscal Effects of Education Finance Reform on Student Achievement**

The shift in fiscal responsibility for education from local school districts to the states has ultimately been driven by a desire to raise the level of student performance, for all or at least some students. Despite the claim by some researchers (e.g. Hanushek, 2003) that changes in the level of spending are unlikely to have much effect on educational outputs, the push for resource equity and adequacy has long rested on the assumption
that shifts in the level or distribution of funding will have real consequences for students’ educational opportunities and success. An important question, then, is whether the fiscal transformation described above has in fact improved student outcomes. The evidence on this question is mixed.

For example, in looking at the relationship between state funding of education and SAT scores, both Peltzman (1993) and Husted and Kenny (1997) found that states with more centralized school funding have lower test scores (or lower test score growth) than decentralized states. Card and Payne (2002), on the other hand, showed that states under court-ordered finance reform saw greater reductions in the test score gap between low and high-income students on the SAT than states with no court-mandated reform. Of course these studies are somewhat limited by their use of college admissions tests relevant only to a select population of students in each state.

Downes and Figlio (1998) showed using two longitudinal surveys of high school seniors from 1972 and 1992 that—controlling for state tax and expenditure limitations that in some cases coincided with school finance reform—court-mandated and legislature-originated reforms improved mean school performance, as measured by test scores, particularly in initially low-spending districts. Tax limitations themselves, however, were shown to have a negative effect on student outcomes (see also Figlio, 1997, and Downes & Figlio, this volume). Both Downes (1997) and Hoxby (2001) considered the impact of school finance equalization on high school dropout rates. Downes (1997) found little to no evidence that state intervention into local school finance improved high school dropout rates, while Hoxby (2001) likewise found no evidence that states with stronger equalization schemes saw greater reductions in dropout rates. One plausible explanation for these findings may be that many finance reforms based on equalization of taxable property wealth do not benefit central city districts, which have historically poor student outcomes but relatively high per-pupil property wealth.

Several papers have assessed the impact of individual state reforms on student outcomes. For example, Downes (1992) found virtually no difference between the distributions of sixth-grade scores on a California state achievement test across school districts before and after Serrano II (1976), despite substantial equalization in per-pupil expenditure. Likewise, Clark (2003) evaluated Kentucky’s Education Reform Act of 1990 that drastically reduced per-pupil spending inequities and found no evidence of a reduction in the achievement gap between high- and low-income students. Flanagan and Murray (2004) also found a small but statistically insignificant effect of increased spending on test scores in Kentucky. Guryan (2001) showed that fourth-grade test scores improved—particularly at the bottom of the test score distribution—after a 1993 reform in Massachusetts, but eighth-grade scores were unchanged. Papke (2005, 2008) found a positive and statistically significant effect of spending equalization on math test pass rates in Michigan after its 1994 reform (see also Cullen & Loeb, 2004; Chaudhary, 2009; and Roy, 2011).

If a lesson can be drawn from the burgeoning research on education finance reform and its effects on student outcomes, it is that these reforms—no matter how far reaching or well intended—are no guarantee of improved adequacy or equity in student achievement. Our review of the literature offers several possible explanations for this. First, even the most generous estimates of the spending-test-score gradient would suggest that extraordinarily large increases in spending are necessary to observe any real increase in student outcomes. It may be that few finance reforms to date have been aggressive enough to show much of an effect on test scores. Second, the institutional details of school funding reforms, the unique circumstances of individual states, and local and state responses to changes in finance systems vary considerably. A firm understanding of these differences is likely to be critical for understanding how a new finance regime will ultimately affect outcomes of the education process. Finally, the ability of a finance reform to improve outcomes may depend on more than just changes to the level and distribution of spending. Arguably, the most successful reforms pay as much attention to the use of funds as to the level of expenditure.28

**Conclusion**

Over the course of the 20th century, states emerged as the predominant source of funding for public K–12 education in the United States, reversing a long-standing tradition of local school finance. This role increased markedly in recent years in the wake of court-ordered and legislative reforms targeting improved equity and adequacy in student resources. As this chapter has demonstrated, the increased state role in funding public
education has indeed affected the level and distribution of school resources within states—on average, states affected by court-mandated finance reforms have increased expenditures and reduced interdistrict disparities to a much greater extent than those states unprompted by litigation. Although there are a few notable exceptions, states in general do not appear to have "leveled down" expenditures in their ongoing efforts to improve equity and adequacy.

The U.S. experiment in greater centralization of school funding remains, however, in its infancy. To date, we have very little evidence of how increased state control over education finance is likely to affect public spending on schools over the long run. Simple models of collective choice would suggest that the dynamics of school spending at the state level are likely to be quite different than those experienced at the local level. Competing state budget interests, greater voter heterogeneity and a more volatile tax base are but three challenges schools are likely to encounter under a more centralized system. The increased involvement of states in funding local schools has in some cases been further accompanied by greater accountability for school performance, or an increased role for school choice, the effects of which are only beginning to be studied. How greater state control over the funding of local schools will ultimately affect the support for public education remains an open and important question.

Notes

1. Some authors treat the terms "reform" and "equalization" in school finance as synonymous. Indeed, early state interventions into local school finance were attempts to equalize differences in local fiscal capacity (through foundation and guaranteed tax base formulas, for example). As the more recent wave of court challenges and legislative reforms have focused less on equalization or "fiscal neutrality," we prefer to use the term "reform" to refer to any significant policy change intended to alter the level and/or distribution of education spending. We refer to "reform states" as any state that has enacted or has been ordered to enact such a change.

2. Hawaii and the District of Columbia are excluded here. Hawaii—with its single school district—had 100 percent of its nonfederal funding classified as state; the District of Columbia had 100 percent of its nonfederal funding classified as local.

3. It may not be immediately obvious how one can reconcile the observation that most states increased their share of K–12 revenues between 1970 and 2000 with the statement that the aggregate state share rose only slightly between 1978 and 2007 (45.6 to 48.3 percent). Part of the explanation lies in timing (the aggregate state share rose 6 percentage points between 1970 and 1978). More importantly, however, is the observation that—in any given year—the aggregate state share is a weighted sum of individual state enrollment shares, where the weights are the relative size of the 50 states in terms of revenue.

4. See Section III of this handbook. Note that concerns over appropriate funding for at-risk students are not exclusively tied to the "adequacy" movement—see, for example, Berne and Stiefel’s (1984) discussion of vertical equity in school finance.

5. These states are Delaware, Hawaii, Mississippi, Nevada, and Utah (ACCESS, 2014).

6. Our list of court rulings has been substantially revised and updated from earlier work (Murray, Evans, & Schwab, 1998; Corcoran et al., 2004) and reconciled against other frequently cited reform inventories (Odden & Picus, 2004; Appendix A of Yinger, 2004; Minorini & Sugarman, 1999; Card & Payne, 2002; Springer, Liu, & Guthrie, 2009; and others). Our list is more conservative in that it includes only rulings made by the highest court in the state explicitly addressing the constitutionality of the school finance system; it also omits rulings that pertain only to capital financing, as we would not expect these rulings to necessarily impact current expenditures. High-court rulings on the justiciability of a dispute related to school financing are not included. We do not deny that such rulings can send strong signals to state legislatures, and in many cases states have responded to these rulings by enacting serious reform. However, in other cases states have not acted until ordered to do so by the state supreme court. For uniformity we consider only orders issued by high courts.

7. See also Campbell and Fischel (1996), Howard and Roch (2001), and Wood and Thoebold (2003) for an alternative view.

8. Exceptions include Evans, Murray, and Schwab (1997) and Downs and Shah (2006). We discuss this literature in greater detail in the following section. Studies of the effects of specific state reforms—whether originating in the legislature or the courts—are more common.


10. We use a modified version of this panel, constructed by Corcoran and Evans (2010).

11. See Berne and Stiefel (1984), Murray, Evans, and Schwab (1998), and Section III of this volume for more on the measures of spending

For reasons explained in Harris, Evans, and Schwab (2001), the sample of school districts upon which these measures are based is somewhat smaller in 1971–72 than in 2006–07. The missing districts in 1971–72 are predominately rural, low-enrollment districts, which suggests (if these districts are relatively low spenders) that we may be understimating inequality in that year. Our calculations are weighted by district enrollment, however, which should alleviate this problem.

These comparisons are similar when other measures of inequality (such as the 95th percentile to 5th percentile ratio) are used in place of the Gini coefficient.

As Silva and Sonstelie (1995) point out, it does not necessarily follow that California’s relative drop in spending was a direct consequence of the Serrano judgment. Most of the loss of school revenue was due to Proposition 13, not Serrano, and while Serrano was an exogenous event, Proposition 13 was an endogenous decision by voters to cap tax revenues. In fact, California had the capacity to increase school funding post–Proposition 13 through other tax sources, but opted not to do so. Therefore, Silva and Sonstelie argue, Serrano fundamentally altered the political determination of education expenditure.


Courant and Loeb (1997) provide details on this reform.

See also Duncombe and Johnston (2004), Downes (2004), and Guryan (2001).

Silva and Sonstelie (1995) refer to this as an "income effect." There may also be a "price effect" if the tax price facing the median voter differs from that of the mean (say, due to the deductibility of local taxes and differing marginal tax rates). The simple model described here also assumes the median voter has the state’s median income.

It is not clear, however, that Texas “leveled down” in the 1990s. While per-pupil spending inequality fell moderately (see Imazeki and Reschovsky (2004), Texas actually ranked fairly high (15th) in real expenditure growth during the 1990s. "Robin Hood" did, however, result in a system of escalating property tax rates that was ultimately found to be unconstitutional (Neeley v. West Orange-Cove Consolidated Independent School District).

The specification of $Dn$ takes a number of forms in this literature. In its simplest form, $Dn$ is an indicator variable that equals one in the year of reform and all subsequent years. Alternative specifications have allowed the effect of reform to vary over time, where $ Dit $ represents the number of years since reform, or is a vector of two or more categorical variables (e.g., 1–5 years since reform, 6 or more years since reform).

Our school district panel is described in Corcoran and Evans (2010). For reasons outlined there, we exclude Alaska, Hawaii, Montana, Vermont, and the District of Columbia, and we restrict our analysis to unified (K–12) districts to control for differences in cost structures across district types. According to the Common Core of Data (2001–02), over 92 percent of all public school students attend unified districts. Nonunified districts are disproportionately located in a few states (e.g., Arizona, California, and Illinois).

Recall that a decrease in each of these inequality measures represents a drop in interdistrict disparities in per-pupil expenditure.

We relied on Odden and Picus (2004), Springer, Liu, and Guthrie (2009), and ACCESS (2014) for the classification of rulings as equity-based or adequacy-based. See Briffault (2007) and Baker and Welner (2011) on the limitations of using a simple categorization of court rulings as strictly "equity" or "adequacy."

Bahl, Sjoquist, and Williams (1990) found among early finance reforms that in nearly all cases increased state aid was offset by reductions in spending inequality implemented here.

The "tax price" implicit in a state aid formula represents the additional tax revenue that must be raised locally in order to increase school expenditure by a dollar. A purely local system of school funding would involve a tax price of $1 in every district. Matching grants from the state can significantly lower the tax price above $1 in high-property wealth districts. See Picus, Goertz, and Odden, this volume, for more details.

A small literature examines how increased state aid is spent. Using national data, Sims (2011b) finds the largest increases in spending following finance reforms goes to teacher salaries, smaller class sizes, and support spending. Salary increases appeared to be more heavily
weighted toward experienced teachers than new ones. In her study of Michigan finance reform, Chaudhary (2009) similarly finds that new aid went to teacher salaries and smaller classes. In Kentucky, Streams et al. (2011) find that state aid provided higher salaries for Appalachian teachers, equalizing pay across Appalachian and non-Appalachian regions (at least in the short run). Finally, in Massachusetts, Dee and Levine (2004) find increased state aid in that state went primarily to instruction and capital investment.

References


## Appendix

### Appendix Table 21.A.1
State Supreme Court Rulings on School Finance System Constitutionality and Pending Cases

<table>
<thead>
<tr>
<th>State</th>
<th>Case</th>
<th>Citation</th>
<th>Year</th>
<th>Ruling</th>
<th>Case Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona</td>
<td>Sheppard v. Yellow</td>
<td>110 Ariz. 88, 512 P.2d 590</td>
<td>1973</td>
<td>Upheld</td>
<td>Equity</td>
</tr>
<tr>
<td>Arkansas</td>
<td>DePue v. Aze School District No. 30</td>
<td>379 Ark. 340, 453 S.W.2d 240</td>
<td>1970</td>
<td>Overturned</td>
<td>Equity</td>
</tr>
<tr>
<td>Iowa</td>
<td>Lake View School District No. 15 v. Blackhawk (Reidville I)</td>
<td>301 Ark. 31, 90 S.W.2d 672</td>
<td>1953</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>Iowa</td>
<td>Lake View School District No. 15 v. Blackhawk (Reidville V)</td>
<td>304 Ark. 398, 250 S.W.2d 645</td>
<td>1953</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>California</td>
<td>Serrano v. Pitzer (Simmons II)</td>
<td>14 Cal.4th 718, 91 P.3d 529, 135 Cal. Rptr. 360</td>
<td>1976</td>
<td>Overturned</td>
<td>Equity</td>
</tr>
<tr>
<td>Colorado</td>
<td>Logan v. Colorado State Board of Education</td>
<td>240 P.3d 188</td>
<td>1983</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Colorado</td>
<td>Lobo v. Colorado</td>
<td>208 P.3d 218</td>
<td>2013</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Connecticut</td>
<td>Burton v. Madoff (Horton I)</td>
<td>127 Conn. 415, 376 A.2d 109</td>
<td>1977</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>Connecticut</td>
<td>Burton v. Madoff (Horton III)</td>
<td>155 Conn. 24, 406 A.2d 1099</td>
<td>1979</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>Florida</td>
<td>Committee for Adjudication and Retrenchment in School Funding v. Hill</td>
<td>869 So.2d 400</td>
<td>1996</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Florida</td>
<td>Committee for Strong Schools v. Florida State Board of Education</td>
<td></td>
<td></td>
<td>Pending</td>
<td>Boyle</td>
</tr>
<tr>
<td>Idaho</td>
<td>Thompson v. Egypt</td>
<td>96 Idaho 703, 527 P.2d 635</td>
<td>1973</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Illinois</td>
<td>Rizzi v. Illinois</td>
<td>185 Ill. 311, 67 N.E.2d 46</td>
<td>1950</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Illinois</td>
<td>Committee for Educational Rights v. Edgar</td>
<td>174 Ill. 2d 1, 602 N.E.2d 1178</td>
<td>1987</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Illinois</td>
<td>Lewis v. Jaspert</td>
<td>180 Ill. 2d 986, 370 N.E.2d 708</td>
<td>1999</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Kansas</td>
<td>Minnery v. State (Moline I)</td>
<td>282 Kan. 9, 134 P.3d 275</td>
<td>2006</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Kentucky</td>
<td>Rowe v. Council for Better Education</td>
<td>790 Ky. 2d 116</td>
<td>1988</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>Maryland</td>
<td>Venable v. Anne Arundel County Board of Education</td>
<td>395 Md. 793, 420 A.2d 788</td>
<td>1963</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Maryland</td>
<td>Brown v. State (Monroe)</td>
<td>278 Md. 522, 373 A.2d 991</td>
<td>1971</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Maryland</td>
<td>Brown v. State (Montgomery)</td>
<td>278 Md. 522, 373 A.2d 991</td>
<td>1971</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Nevada</td>
<td>Harrell v. Board of Education (Board of Education)</td>
<td>88 Nev. 240, 447 P.2d 1185</td>
<td>1970</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>Nevada</td>
<td>McKee v. Board of Education (Board of Education)</td>
<td>102 Nev. 269, 224 N.W.2d 571</td>
<td>1971</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>Minnesota</td>
<td>Skoog v. Minnesota (1)</td>
<td>505 Minn. 229</td>
<td>1953</td>
<td>Upheld</td>
<td>Boyle</td>
</tr>
<tr>
<td>New York</td>
<td>Committee for Educational Equality v. Monroe (I)</td>
<td>604 N.Y.2d 279</td>
<td>1990</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>New York</td>
<td>Committee for Educational Equality v. Monroe (II)</td>
<td>690 N.Y.2d 57</td>
<td>1992</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>New York</td>
<td>Board v. Board of Education (Board of Education)</td>
<td>142 N.Y. 402, 703 A.2d 155</td>
<td>1997</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>New York</td>
<td>Committee for the Justice-—School Finance (Boylston II)</td>
<td>142 N.Y. 402, 703 A.2d 155</td>
<td>1997</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>New York</td>
<td>Opinion of the Justice—School Finance (Boylston II)</td>
<td>142 N.Y. 402, 703 A.2d 155</td>
<td>1997</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>New York</td>
<td>Opinion of the Justice—School Finance (Boylston V)</td>
<td>142 N.Y. 402, 703 A.2d 155</td>
<td>1997</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>New York</td>
<td>Opinion of the Justice—School Finance (Boylston V)</td>
<td>142 N.Y. 402, 703 A.2d 155</td>
<td>1997</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>New York</td>
<td>Opinion of the Justice—School Finance (Boylston V)</td>
<td>142 N.Y. 402, 703 A.2d 155</td>
<td>1997</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
<tr>
<td>New York</td>
<td>Opinion of the Justice—School Finance (Boylston V)</td>
<td>142 N.Y. 402, 703 A.2d 155</td>
<td>1997</td>
<td>Overturned</td>
<td>Boyle</td>
</tr>
</tbody>
</table>
Albert v. Becker (Albert II) 399 N.J. 140 2000 Upheld Adequacy
Albert v. Becker (Albert III) 206 N.J. 332 2011 Overturned Adequacy

New York
Lee v. New York State 315 N.Y. 2d 1, 494 N.E.2d 319. 1982 Upheld Equity
Reid v. New York State Education Department (Rigler v. Genaro) 86 N.Y.2d 376, 635 N.E.2d 647. 1994 Upheld Equity
Campbell v. New York State Board of Regents 98 N.Y.2d 499 2002 Overturned Adequacy
Albert v. State Pending Adequacy

North Carolina
Mallor v. State Pending Adequacy

Mike County Board of Education v. North Carolina (Conard II) 315 N.C. 2d 247 1994 Upheld Equity

North Dakota
Reinbold v. North Dakota Board of Education 205 N.D. 2d 370, 558 N.W.2d 24 1997 Upheld Equity
Delbridge v. State (Delbridge I) 161 N.D. 2d 367, 697 N.W.2d 488 2005 Upheld Adequacy
Delbridge v. State (Delbridge II) 58 N.D. 2d 367, 751 N.W.2d 486 2005 Upheld Adequacy
Delbridge v. State (Delbridge III) 58 N.D. 2d 367, 751 N.W.2d 486 2005 Upheld Adequacy
Delbridge v. State (Delbridge IV) 58 N.D. 2d 367, 751 N.W.2d 486 2005 Upheld Adequacy

Ohio
Board of Education v. Shreve 39 Ohio St. 3d 249 1980 Upheld Equity
Delbridge v. State (Delbridge I) 58 Ohio St. 3d 367, 751 N.W.2d 486 2005 Upheld Adequacy
Delbridge v. State (Delbridge II) 58 Ohio St. 3d 367, 751 N.W.2d 486 2005 Upheld Adequacy
Delbridge v. State (Delbridge III) 58 Ohio St. 3d 367, 751 N.W.2d 486 2005 Upheld Adequacy

Pennsylvania

Rhode Island
City of Warwick v. School Board 463 A.2d 46 1983 Upheld Adequacy
Warren v. Warwick Pending Adequacy

South Carolina
Exeter County v. Campbell 294 S.C. 364, 364 S.E.2d 879 1988 Upheld Equity
Allendale County School District v. South Carolina Pending Adequacy

South Dakota
Dietz v. State 288 N.W.2d 518 1980 Overturned Adequacy

Tennessee
Tennessee School Finance Courts v. McWhorter (I) 815 S.W.2d 130 1991 Upheld Equity
Tennessee School Finance Courts v. McWhorter (II) 899 S.W.2d 724 1995 Upheld Adequacy
Tennessee School Finance Courts v. McWhorter (III) 91 S.W.2d 358 2002 Upheld Equity

Texas
Edgeuwood v. Independent School District of East Creative 244 S.W.2d 240 1960 Upheld Equity
McFarland v. Independent School District of East Creative 244 S.W.2d 240 1960 Upheld Equity
Terrell v. Independent School District of East Creative 244 S.W.2d 240 1960 Upheld Equity

Utah
Richardson v. Salt Lake School District 165 S.W.2d 847 1942 Upheld Equity

Vermont
Rigler v. Vermont (Rigler I) 36 Vt. 175, 116 Vt. 2d 381 1997 Upheld Equity

Virginia

Washington
South Kitsap School District No. 1 v. Washington 90 Wash. 2d 203, 575 P.2d 711 1978 Upheld Equity
McCook v. Washington 240 Wash. 2d 377 2012 Upheld Adequacy

West Virginia
Pender v. Barley 171 W. Va. 167, 321 S.E.2d 1 1984 Upheld Equity

Wisconsin
Kohler v. Grauer 116 Wis. 2d 463, 346 N.W.2d 568 1984 Upheld Equity
Vance v. Wright 254 Wis. 2d 102, 648 N.W.2d 201 2003 Upheld Adequacy

Wyoming
Pine River School District No 1 v. Natrona 606 Wyo. 3d 310, 399 N.W.2d 101 1980 Upheld Equity
Campbell County School District No. 1 v. Wyoming 957 Wyo. 2d 132, 627 N.W.2d 3 1995 Upheld Adequacy
Campbell County School District No. 1 v. Wyoming (Campbell I) 957 Wyo. 2d 132, 627 N.W.2d 3 1995 Upheld Adequacy
Campbell County School District No. 1 v. Wyoming (Campbell II) 957 Wyo. 2d 132, 627 N.W.2d 3 1995 Upheld Adequacy
Campbell County School District No. 1 v. Wyoming (Campbell III) 957 Wyo. 2d 132, 627 N.W.2d 3 1995 Upheld Adequacy


Notes:
1) Does not include ratings related to capital or facility financing (e.g., Avonlea v. Tennessee (Avonlea I) 410 U.S. 242, 2000 Upheld Equity
2) Does not include high court-generated ratings (e.g., a high court decision of a lower court's dismissal of this suit, assuming the suit is not tried.
3) Does not include intermediate court ratings (e.g., Seneca v. Kentucky (Seneca I) 300 N.E.2d 530 1965 Upheld Equity

326
Local Funding of Schools

The Property Tax and Its Alternatives

THERESE J. McGUIRE, LESLIE E. PAPKE, AND ANDREW RESCHOVSKY

Introduction

The property tax has long been the primary local source of funding for schools and, along with state aid, provides the lion’s share of total resources for schools. In recent decades, though, the property tax has come under siege as a source of revenue for schools. The principal charge has been that, since property wealth is unequally distributed across school districts, reliance on this source of revenue results in unacceptable differences in property tax rates, property revenues per pupil and, most importantly, spending per pupil across districts. The property tax is also criticized for being inequitable, inefficient, and complex. Dissatisfaction with the tax has grown as evidenced by the number of attempts to constrain spending and limit access to taxes, including the property tax (Downes & Figlio, this volume). Among the possible explanations for the declining support for property taxes are demographic shifts in the form of a rising elderly share and a declining school-age share of the population. In this chapter, we examine the various charges against the property tax as a means of funding schools and compare it to alternative sources of revenue such as sales and income taxes.

The Role of the Property Tax in Funding U.S. Education

In 2010–11, the most recent year of available data, the federal government provided 12.3 percent, states provided 44.4 percent, and local governments provided 43.3 percent of public school system revenue. There was, however, substantial variation across states in the percent of revenue provided by local governments. Local governments provided over half of total revenue in 12 states and the District of Columbia, while in 19 states, state governments provided over half of the revenue. In the majority of the remaining 19 states, K–12 funding was relatively evenly shared.

The property tax has been and remains the predominant local source of revenue for school districts. Not all school districts, however, have independent authority to raise revenue. In 36 states, districts have the power to generate their own revenue by setting property tax rates. In seven states plus the District of Columbia, school districts are dependent for revenue on a city, town, or county. Other states have a mix of both dependent and independent school districts. Dependent school systems receive most of their local revenue from appropriations by their parent government. Most of this revenue comes from property tax collections.

Table 22.1 lists sources of local school district revenue in fiscal year 2011. Nearly 66 percent of local revenue comes from the property tax. This figure underestimates the importance of the property tax, however, because much of the 18.1 percent of local revenue from parent government contributions also comes from the property tax. In 2010/11, the property taxes collected by independent school districts provided over 80 percent of local revenues in 20 states and over 50 percent of revenues in 38 states. Only six states received more than 5 percent of local revenues from other taxes. For example, local school districts in Georgia make extensive use of the sales tax.

Comprehensive annual data on school property tax revenue in the United States are only available for the years since 1988/89, when the National Center for Education Statistics (NCES) began collecting these data as
part of their National Public Education Financial Survey of state education agencies. In order for us to present data on school property tax revenues for the years prior to 1989, we must combine data on the property tax revenue of independent school districts from the Census of Governments with an estimate of the property tax component of parent government contributions to the total revenues of dependent school districts. Unfortunately, these calculations are not possible for years before 1977 because in all earlier years the Census of Governments only collected revenue data from independent school districts. In the next paragraph, we describe the steps we follow to estimate public school property tax revenue for the years between 1977 and 1989.

Table 22.1 Percent Composition of Local Source School District Revenue, 2010/11

<table>
<thead>
<tr>
<th>Source of Revenue</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property taxes</td>
<td>65.6</td>
</tr>
<tr>
<td>Other taxes</td>
<td>3.1</td>
</tr>
<tr>
<td>Parent government contribution</td>
<td>18.1</td>
</tr>
<tr>
<td>Nonschool local government</td>
<td>2.5</td>
</tr>
<tr>
<td>School lunch charges</td>
<td>2.5</td>
</tr>
<tr>
<td>Tuition and transportation charges</td>
<td>0.5</td>
</tr>
<tr>
<td>Other charges</td>
<td>2.5</td>
</tr>
<tr>
<td>Other local revenue</td>
<td>6.2</td>
</tr>
</tbody>
</table>

Source: U.S. Census Bureau, 2013.

Using data from the 1989 National Public Education Financial Survey, we calculate that 75.3 percent of parent government contributions to dependent school districts came from the property tax. Census data on the financing of municipal and county governments during the 1970s and 1980s indicates that over this period municipal and county governments reduced their reliance on the property tax at the rate of about 1.25 percent per year. Using this information, we calculate the share of total parental government contributions attributable to the property tax for the years these data are available from the Census of Governments, namely 1977, 1981, 1982, and 1987. For each of these years, we then calculate property tax revenues from both independent and dependent school districts as a share of total local government education revenues, and then linearly interpolate these shares for the intercensal years between 1977 and 1989. Total property tax revenues for these years are calculated by multiplying these property tax shares by total local education revenue in each year as reported in the NCES Digest of Education Statistics.

The results of these calculations appear in Table 22.2, which reports the percent of K–12 revenue from federal, state, and local sources, and from the property tax for the years 1976/77 to 2010/11. Over this period, per student school property taxes in the United States grew from $2,547 to $4,264 (in 2011 dollars), an average annual rate of growth of 1.5 percent. Given the widespread dissatisfaction with the property tax, it is perhaps surprising how stable the property tax has been as a source of education revenue. As illustrated in the last column of Table 22.2, after declining by 5 percentage points between 1976/77 and 1981/82, the share of total revenue from the property tax remained quite constant at about 35 percent. The initial decline in the role of the property tax reflects both an increase in state governments’ role in education funding that occurred between the mid-1970s and the mid-1980s and a reduction in the share of local government revenue coming from the property tax. Although not shown in the table, the share of local government revenues from property taxes fell from 83.0 percent in 1977 to 77.1 percent in 1982.

The role of the property tax in school funding, while quite stable since the early 1980s, declined sharply in previous decades. Although comprehensive data on school property tax revenues are unavailable for earlier years, school property tax revenues can be approximated from the fact that most local government funding of public schools comes from the property tax and the share of education revenue from local government sources declined sharply through most of the 20th century. Prior to 1930, local governments contributed over 80 percent of school revenues. The local government share fell to 68 percent by 1940 and to 57.3 percent by 1950. The local share remained above 60 percent through 1974, declining to 45 percent in 1982. As shown in Table 22.2, the local share has remained largely unchanged since then.

Table 22.2 Sources of School District Funding: Selected Years 1976/77–2010/11

329
Currently, the local property tax comprises 35 percent of the revenue of K–12 school districts. This figure understates the importance of the property tax, however, because it includes only local property tax revenue. In several states, statewide property taxes have replaced or supplemented the local property tax. California and Michigan, for example, have both capped the ability of their school districts to use the property tax and have in effect adopted statewide property taxes to fund K–12 schools. In other cases, including, for example, New Hampshire, states have implemented a foundation formula with a required minimum tax rate contribution by local school districts that is labeled as a statewide property tax. A number of states have enacted property tax limits that restrict the amount of additional property tax revenue that can be raised each year. For example, Proposition 2½ in Massachusetts limits annual increases in property tax levies to 2.5 percent a year. An implication of moving to a state property tax or of putting limits on revenue from local property taxes is that the property tax is no longer the marginal source of funds for local school districts. Generally, there are efficiency gains when the last dollar of school funding comes from a local tax—the property tax in most states—so that spending decisions can reflect local preferences for education. Thus, the shift to statewide property taxes or the imposition of limits on local property tax revenue could lead to a less efficient distribution of spending across districts.

In 1994, Michigan replaced local property taxes as the primary funding source for K–12 funding with a combination of a flat-rate state property tax and a portion of the state sales tax. Prior to the reform, in 1992, local property taxes accounted for 55.9 percent of K–12 funding. The property tax share fell to 22.0 percent in 1997, and rose to 28.2 percent in 2011. The state-level property tax for education in Michigan is levied at six mills on both homestead and nonhomestead property. Proposal A, as the funding reform was called, established a minimum per-pupil foundation grant and imposed a cap on property tax revenue. These changes resulted in an overall net drop in property taxation. In 1993, the average statewide millage rate for all purposes was 56.6 mills. In 2000, a few years after the passage of Proposal A, the statewide average homestead millage rate had dropped to 31.5 mills, and the nonhomestead rate to 50.1 mills.

California’s Proposition 13, which passed by voter initiative in 1978, limited local governments’ ability to fund education by establishing a maximum local property tax rate of 1 percent, rolling back assessments to 1975 levels and limiting reassessments to an increase of 2 percent per year, except upon the sale of a property. The property tax revenues do not stay in the local jurisdiction where they are raised. Rather, the revenues from the uniform 1 percent property tax are disbursed by the state through grant programs to local jurisdictions. These changes essentially converted the local property tax into a state property tax. In 1957, the property tax comprised 48.9 percent of total revenues and remained about 50 percent of revenue through 1977, but by 1982 it had fallen to 21.5 percent of total revenues. In 2011, property tax revenues stood at 23.7 percent of revenues, local funding (which includes property tax revenues) was 29.8 percent, and state funding, 56.6 percent.

Similarly, school funding reforms in New Hampshire and Oregon resulted in reductions in the share of public school revenue from the local property tax of over 25 percentage points, and the role of the state

<table>
<thead>
<tr>
<th>Year</th>
<th>From Federal Sources</th>
<th>From State Sources</th>
<th>From Local Sources</th>
<th>From Local Property Taxes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976/77</td>
<td>8.8%</td>
<td>43.4%</td>
<td>47.8%</td>
<td>39.7%</td>
</tr>
<tr>
<td>1981/82</td>
<td>7.4%</td>
<td>47.6%</td>
<td>45.0%</td>
<td>34.7%</td>
</tr>
<tr>
<td>1982/87</td>
<td>6.4%</td>
<td>49.7%</td>
<td>43.9%</td>
<td>35.5%</td>
</tr>
<tr>
<td>1988/89</td>
<td>6.2%</td>
<td>47.8%</td>
<td>45.6%</td>
<td>35.8%</td>
</tr>
<tr>
<td>1991/92</td>
<td>6.6%</td>
<td>46.4%</td>
<td>46.6%</td>
<td>36.6%</td>
</tr>
<tr>
<td>1996/97</td>
<td>6.6%</td>
<td>48.0%</td>
<td>45.0%</td>
<td>34.9%</td>
</tr>
<tr>
<td>1999/2000</td>
<td>7.3%</td>
<td>49.5%</td>
<td>42.9%</td>
<td>33.4%</td>
</tr>
<tr>
<td>2004/05</td>
<td>9.2%</td>
<td>46.9%</td>
<td>43.6%</td>
<td>34.4%</td>
</tr>
<tr>
<td>2005/06</td>
<td>9.1%</td>
<td>46.5%</td>
<td>44.0%</td>
<td>34.2%</td>
</tr>
<tr>
<td>2006/07</td>
<td>8.5%</td>
<td>47.4%</td>
<td>43.8%</td>
<td>33.9%</td>
</tr>
<tr>
<td>2007/08</td>
<td>8.2%</td>
<td>48.3%</td>
<td>43.3%</td>
<td>33.6%</td>
</tr>
<tr>
<td>2008/09</td>
<td>9.6%</td>
<td>46.7%</td>
<td>43.5%</td>
<td>34.7%</td>
</tr>
<tr>
<td>2009/10</td>
<td>12.7%</td>
<td>43.5%</td>
<td>43.5%</td>
<td>35.3%</td>
</tr>
<tr>
<td>2010/11</td>
<td>12.5%</td>
<td>44.1%</td>
<td>43.1%</td>
<td>35.0%</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations using data from the National Center for Education Statistics and the U.S. Census Bureau. See text for details.
government in school funding was dramatically increased. In Vermont, the responsibility for the funding of public education was almost entirely shifted from local governments to the state government. Although local school property tax funding was essentially eliminated, a new state property tax was adopted.

During the same period, however, local property taxes as a share of total school revenue increased in a number of states. For example, in Florida, Kentucky, Missouri, Nevada, Pennsylvania, and Washington, the share of revenue from the property tax rose by more than 5 percentage points. In each of these states, the state government’s role in school funding was smaller in 2009–10 than it had been in 1989–90.

The role of the property tax in the funding of public education varies dramatically across states. In 2009–10 in five states (Connecticut, Illinois, New Hampshire, New Jersey, and Rhode Island) the property tax contributed over half of public education revenue. In sharp contrast, in Alabama, Alaska, and New Mexico less than 15 percent of public school revenues came from the property tax. Not only did these three states rely heavily on state funding for education, but local governments in these states raised a substantial portion of the revenues they devoted to education from the sales tax.

**Implications of 40 Years of Court Challenges and Legislative Actions**

Many charges have been made against the property tax to justify actions that reduce reliance on the tax for school funding. Critics have argued, for example, that the property tax is regressive, that it distorts economic behavior, that it is difficult to administer, and that its burden is unrelated to the ability of taxpayers to pay. But these charges—some valid, others not so valid, as we argue below—pertain to the property tax in general, whether the tax is used to fund education or parks.

Another charge—that disparities in property wealth lead to disparities in spending—is applicable to any type of local government expenditure but is particularly relevant to local schools. It is this aspect of the local property tax that has motivated school finance court cases. Plaintiffs have used provisions in state constitutions for equal protection and fair and adequate provision of education to argue for reduced reliance on local property taxes for funding schools. None of the legal arguments employed specifically condemns the property tax as a means of financing education. Rather, it was the use of the property tax by local jurisdictions, whose per-pupil property tax bases varied considerably, that was deemed problematic. From the arguments made it would seem that any locally imposed tax whose base varied greatly across jurisdictions would have been challenged.

Legal challenges brought under equal protection clauses of state constitutions attempted to reduce the interdistrict variation in per-pupil spending and to weaken the link between local property wealth and school spending. California is widely viewed as having gone the furthest in this respect with the changes implemented in response to the two California Supreme Court decisions related to school funding (Serrano I and II) in the early to mid-1970s and the subsequent passage in 1978 of the tax limitation measure, Proposition 13. As discussed above, local school districts (and other local taxing bodies) continue to derive significant revenues from property taxes in California, but the tax has effectively become a state tax—imposed at a uniform rate of 1 percent.

Another set of legal challenges has been brought under education clauses of state constitutions. The argument in most of these cases has shifted away from the concept of spending equity with its focus on the variation in per-pupil spending across districts or on fiscal neutrality with its focus on the relationship between property tax wealth and district spending toward the newer standard of educational adequacy. Under this standard, the state attempts to ensure that every student has access to an adequate level of education. (See Koski & Hahnel and the chapters in Section III, this volume.) Thus, state policy attention shifts almost exclusively to the low-spending districts.

The shift in emphasis from equity to adequacy could have fairly dramatic implications for the property tax. The California case is illustrative. The system the state implemented in response to the first Serrano decision attempted to reduce interdistrict disparities in spending per pupil by raising the spending of low-spending districts through a stronger foundation aid program and limiting the spending of high-spending districts through a limit on local property taxes. In Serrano II, the State Supreme Court ruled that this new system continued to violate the equal protection clause of the constitution because it continued to allow districts to
override the limit on property taxes. If the system had been challenged under an adequacy standard instead, binding limits on property taxes might not have been necessary and breaking the link between local property wealth and school spending that the Court demanded, and that Fischel (1989) argues was central to the support for Proposition 13, might not have occurred.\textsuperscript{11}

Using data from 1972 to 2002 on per-pupil spending across states, Corcoran and Evans (this volume) estimate the average impact of court-ordered reforms on per-pupil spending across states. They find that (1) court-ordered reforms resulted in a decline in within-state inequality in spending per pupil of between 15 and 19 percent over this period; and (2) after court-ordered reform, spending rose 12.5 percent more in the low-spending school districts than in comparable ones in nonreform states, and remained essentially constant in the highest-spending districts. They present evidence suggesting that court rulings on the basis of adequacy have had less impact on interdistrict spending disparities than those based on equity arguments.

Missing from previous research is any direct evidence on how court-ordered reform has affected the level or structure of property taxes. Using calculations provided to us by Kim Rueben (personal communication), we compare changes in the percent of elementary and secondary expenditures financed by property taxes in states subject to court-ordered reform to the corresponding figures in states without court-ordered reform.\textsuperscript{12} We find that the property tax is diminished in all states, whether subject to court-ordered reform or not, but it appears that the decline in the role of the property tax may have been somewhat greater in states subject to court-ordered reform. For states with post-1985 court orders, the share of expenditures financed by the property tax fell from 38.0 percent in 1992 to 30.6 percent in 2002, a drop of nearly 7.5 percentage points in 10 years. The drop for states without court orders was less than 5 percentage points (from 36.2 in 1992 to 31.7 percent in 2002).

The shift in emphasis in school-reform court challenges from equity and fiscal neutrality to adequacy may hold out hope for the property tax. While the emphasis on equity and fiscal neutrality generally has led to less reliance on local property taxes, how the current emphasis on adequacy will affect property taxes is not obvious. On the one hand, because adequacy is likely to require an increase in total spending on education, the shift may lead to an even greater role for the state in funding education. At the same time, though, it need not lead to a reduced reliance on property taxes. Adequate funding could potentially be achieved through the use of a (perhaps, significantly enhanced) state foundation aid program in conjunction with a fully functioning local property tax. Under some interpretations of adequacy there would be no need to limit local property taxes or to restrain spending by high-wealth districts.\textsuperscript{13} As long as the foundation level was set at a high enough level to achieve adequacy and the level was guaranteed with funding from the state, the goal of providing access to an adequate level of education could be achieved without eliminating or severely crippling the local property tax. In other words, if districts chose to spend more than the state guarantee, they could do so by tapping into their property tax wealth, and this behavior would not compromise the adequacy goal. Of course, policymakers may wish to pursue both adequacy and equity, in which case limits on the property tax, additional equalizing state aid programs, or restrictions on the supplements that districts can add to the foundation level may be desirable in order to keep high-wealth districts from spending more than low-wealth districts.\textsuperscript{14}

Is the Property Tax Structurally Flawed?

Some critics of school funding systems focus on the very nature or structure of the property tax itself—that it is inefficient, inequitable (in that the incidence falls disproportionately on lower-income taxpayers), and difficult to administer. These criticisms are not specific to the funding of school districts, and, in large part, they pertain to the property tax whether it is a local or a state tax.

The property tax consists of many taxes. It may be a tax on land, a tax on improvements to land, including residential buildings and commercial and industrial buildings, and a tax on personal property, including tangible personal property such as furniture, machinery and equipment, and intangible personal property such as stocks and bonds. The property tax is levied on property as valued by assessors, and placing values on many of the components of the property tax base can be difficult. For example, one acre of land in the center of a
large city can be more or less valuable depending on what the parcel of land is proximate to (a port? a large condominium building? a public housing project?), the history of the use of the parcel (is it contaminated with toxic materials?), and the predictions for the future economic health of the neighborhood. It is a tax on capital wealth, and yet property represents consumption of housing and other services. The particular structure of the property tax has implications for the efficiency, equity, and simplicity of the tax.

The charge that the tax is inefficient refers to distortions in decisions regarding factors of production, such as capital, that can be moved either across jurisdictions or among sectors of the economy in response to property tax rates. Zodrow (2001) summarizes the basic argument that the use of property taxes to finance local public services results in inefficiently low levels of public services, as jurisdictions try to keep the property tax low in an effort to keep mobile capital, such as manufacturing firms, from moving to other lower-tax jurisdictions. Competition of this form for mobile capital is inefficient because it creates a fiscal externality—when a given jurisdiction increases its tax rate and drives capital out, it creates a fiscal benefit for the jurisdictions to which the capital flows (see Wildasin, 1989). Still, Inman (2001) argues that the property tax is likely to be no more inefficient than other local taxes that could serve as alternative sources of revenue. He argues that mobile workers, capital, and consumers (but not land) can escape any local tax by moving to other cities. If these other cities are similar in the efficiency of their governments and economies, which may be plausible in suburban America, then there will be little loss in social welfare associated with a local property tax.

An alternative argument is that the property tax is a benefit tax and therefore is an efficient and nondistortionary source of local revenue (Fischel, 2001). This argument rests on the assumption that there is a tight link between taxes paid and services received, an argument that is quite plausible for services such as police and fire protection. With respect to schools, especially inner-city schools, however, the argument is difficult to make because the value of school services received is likely to have little relation to housing value and therefore to a taxpayer’s property tax liability. While competition among school districts in the suburbs can lead to capitalization of school services quality into housing values, thereby providing the requisite link for the property tax to act as a benefit tax, in school districts in large central cities, which serve a large percentage of public school children, competition is usually lacking. Thus, the argument that the property tax acts like a benefit tax when used to fund education is not very persuasive.

The question of the distributional effects of the property tax, its incidence, has long been the subject of debate among scholars. The traditional view (see Netzer, 1966) argues that the tax is similar to an excise tax on housing, and, because expenditures on housing fall as a percentage of income as income rises, the tax burden is distributed regressively. Beginning with Mieszkowski (1972), scholars began to argue for a "new view" of the burden of the property tax. As the property tax is widely used across the country and applies to land and structures, they argued that the property tax functions nationally like a tax on capital. The fact that capital is in relatively fixed supply in the aggregate means that the property tax serves to reduce the overall average net-of-tax return to capital. Hence, under this view, the tax burden is progressive since high-income individuals receive higher proportions of their income from capital than do lower-income individuals. The observation that there is not one monolithic national property tax and that local tax rates differ, modifies the conclusion only somewhat. The local tax differentials will be borne by landowners and nonmobile consumers and workers. To the extent that high-income taxpayers tend to live in high-tax-rate communities, the burden of the local proportion of the property tax would also be progressive.

These theoretical conclusions rest on some strong assumptions regarding relative supply and demand elasticities and the corresponding ability to shift taxes. The question of the distribution of the burden of the property tax is ultimately an empirical one. Conclusions about the distribution of the tax burden also depend on the definition of income used to calculate tax burdens. Although frequently used in empirical studies, many economists would agree that the use of annual income in calculating tax burdens biases the results towards regressivity. Metcalf (1994) and Poterba (1989) use annual consumption expenditures as a proxy for lifetime income to examine the incidence of various taxes, including the property tax. By this alternative income measure, they find that the burden of the residential property tax is close to proportional. Chernick and Reschovsky (1993) challenge Metcalf’s measure of permanent income on the ground that annual consumption expenditures are or can be highly variable as unexpected and non-recurring expenditures arise. Chernick and Reschovsky substitute 11-year averages for both income and tax payments for measures based on a single year
and conclude that the tax is mildly regressive. Plummer (2003) examines the incidence of residential property
taxes in Dallas County, Texas across thousands of owner-occupied homes. Arguing that individuals rely on
expectations of their lifetime earnings when purchasing housing, she uses residential property value as a
measure of a household’s lifetime income and finds the burden of the property tax to be distributed essentially
proportionally.

Despite the view of many economists that the property tax burden is not distributed regressively, the
layman’s view is largely fixed on the tax being regressive. In part, this view reflects aspects of the tax that have
little to do with determining incidence, but strike people as unfair attributes of the tax: the fact that the tax is
paid in lumpy increments and that tax liabilities do not closely track current incomes. Still, the notion that the
tax is regressive influences public policy. Nearly all states have instituted various exemptions and credits to try
to relieve the burden of the tax on the elderly (with the argument being that the elderly are on fixed incomes)
and on low-income taxpayers. Also, the argument that the tax is regressive has been used in some states to
justify the enactment of various property tax limitations.

A third view of the incidence of the tax is that to the extent that the tax can be viewed as a benefit tax—that
is, a price paid for services received—issues of redistribution and incidence are irrelevant. Fischel (2001) is a
strong proponent of this view. For the property tax to act as a benefit tax several strong conditions must hold,
including free mobility of households among many communities differentiated by their tax and public service
packages. These conditions might accurately describe the suburbs of large metropolitan areas, but they are not
likely to hold in central cities or rural areas.

The property tax is open to criticism because it is difficult to administer fairly. In some jurisdictions,
assessment of property value for taxing purposes may include measurement error, special treatment for certain
classes of taxpayers (such as the elderly), and the possibility of political influence. Assessing property in large
cities is a daunting and potentially costly task, especially when housing prices change rapidly. In addition, it is
difficult to determine the value of property when sales of similar property are infrequent or, as is often the case
for nonresidential property, when no similar properties exist. Consider in contrast, the income tax for which
individual taxpayers have frequent observations and reports on the income on which to base their calculations
of liability. Similarly, with the sales tax, taxes are easily calculated at the time of purchase. In comparison, the
property tax may appear more arbitrary and clumsy. Further, multiple overlapping jurisdictions generate the
aggregate statutory tax rate on a given parcel. The presence of overlapping jurisdictions makes it difficult for
taxpayers to know how much tax they are paying to which governments. And the efforts to relieve the burden
on select taxpayers or to redress perceived inequities and inefficiencies lead to complex exemptions, credits,
and limitations measures. Although most taxes have problems with administrative complexity (for example,
most general sales taxes have exemptions), the property tax seems more complex and less transparent than
income or sales taxes.

Are Demographic and Economic Trends Undermining the Property Tax?

Even if economists generally agree with Wallace Oates (2001, p. 29) that the property tax is “well suited for use
at the local level,” there is no denying that the tax is unpopular and is being increasingly crippled by the
enactment of various limits and restrictions. In the modern era, the passage of Proposition 13 in California in
1978 started a wave of similar limitations and restrictions in states as diverse as Arizona and Massachusetts.
These measures, enacted in 30 states, limit property tax rates, property tax revenues, and property tax
assessments (Downes & Figlio, this volume). The cumulative impact of these limits and restrictions is to render
the property tax (and other taxes subject to the limits) ineffective as a means for providing local governments
with fiscal autonomy.

Current demographic and policy trends may add to the difficulties with the tax. The aging of the U.S.
population, for example, could well reduce support for the property tax. Menchik (2004) documents that the
percentage of the population below the age of 19 is expected to decline from 28.6 percent in 2000 to 26.3 percent
in 2035. Perhaps more importantly, the percentage of the population aged 65 or over is expected to increase
from 12.4 to 20.5 percent over the same period.

As they move away from their children and grandchildren to retirement homes and come to rely on
relatively fixed sources of income, the elderly may be less willing to support high property taxes for schools. Poterba (1997) finds support for this notion. Using state-level panel data for the period 1960 to 1990, he finds that per child spending is negatively related to the elderly share of the population. In a study that challenges this result, Ladd and Murray (2001) use data on county-level expenditures on education and find that differences in the share of county population that is age 65 or older has no impact on the level of education spending per student.

Harris, Evans, and Schwab (2001) use a national panel of individual school districts to study the effect of an aging America on education spending. They find that while the elderly have only a modest negative effect on education spending at the district level, a large share of elderly depresses education spending at the state level. They interpret these results to be consistent with the notion that differences in local, but not state, spending are capitalized into housing values, which homeowners, including the elderly, presumably care about. Brunner and Balsdon (2004) examine evidence from a survey of potential voters on two initiatives on public school spending in California. They find that support for school spending declines with age, and that older voters are more willing to support local than state spending on schools.

Berkman and Plutzer (2004) found that the support of the elderly for local public education spending depends on how long they have resided in a community, with new elderly residents less likely to support public education spending than are long-standing residents. Fletcher and Kenny (2008) using a median voter framework find that the elderly are associated with only a very small drop in the support of education. In a recent article, however, Figlio and Fletcher (2012) develop a method to isolate the impact of aging in place, and find as people become elderly they reduce their support for public education.

Another potentially relevant development is that for many taxpayers subject to the Alternative Minimum Tax (AMT), larger property tax deductions do little or nothing to reduce federal income tax liabilities. While the property tax deduction reduces regular income taxes, it also results in an offsetting increase in the AMT. The AMT thus raises the marginal price of an extra dollar of local government spending funded with the property tax (including education spending) from as low as $0.60 to one full dollar. Without the federal discount provided to the property tax, political support for the property tax (and other deductible state and local taxes) may be diminished.

Alternatives to the Property Tax for Funding Schools

Suppose policymakers are persuaded that the property tax is structurally flawed or that it is too hampered by restrictions to be useful for school funding. Where would they turn for revenues to support schools? Would they turn to a local income or sales tax? If so, how would they address the problems of these other local sources? Would they move all funding of local schools to the state level? If so, how would they address the problems associated with breaking the link between spending responsibility and revenue-raising authority?

McGuire (2001) begins her analysis of possible alternatives to the local property tax by asking whether local governments need a local source of revenue. She notes that local governments in Europe are largely financed by the central government and that in many U.S. states local governments are heavily reliant on state grants because their access to local revenues is severely restricted. Yet, the arguments for fiscally empowering local governments, including schools—at least fiscally empowering them at the margin—are compelling: the ability to cater to differences in tastes for local public goods is enhanced when local governments have authority to raise their own revenues. In addition, when local governments are grant financed the incentive to spend monies responsibly is weakened, especially when grants fund the marginal dollar of spending.

Assuming that a local source of revenues is desired, McGuire compares the local property tax to its two main alternatives—a local sales tax and a local income tax—and argues that the property tax compares favorably along efficiency and equity criteria. In particular, differences across local jurisdictions in sales and income taxes distort purchases and employment decisions perhaps as much as property tax differences distort investment decisions. The typical sales tax is considerably more regressive than the property tax, and while income taxes can be more progressive than property taxes, subnational income taxes are usually nearly proportional or only modestly progressive by design.
Oates and Schwab (2004) set up a horse race between the property tax and the income tax as the primary source of revenues for local governments (they do not focus on schools). On the standard economic criteria of equity and efficiency, neither tax gets the nod over the other. Both taxes are likely to distort economic decisions, but either tax is likely to work well as a signal to mobile residents choosing among local jurisdictions. Although income taxes can be designed to be more progressive than property taxes, in practice they rarely are at the state and local level. Also, under the new view that the property tax is largely a tax borne by owners of capital or under the view that lifetime income should be used in assessing the incidence of the property tax, the property tax compares favorably. On the question of visibility or transparency of the tax, they argue that the property tax—largely because there is only one level of government relying on the tax—is superior to the income tax.

Reschovsky (2004) challenges the argument that the property tax is more transparent than the income tax. He points out that while local governments are the primary users of the property tax, the tax is still complicated because of its use by overlapping layers of local governments such as cities, counties, and schools districts. Reschovsky also argues that the income tax has advantages over the property tax as a local revenue source. Depending on the design of the two taxes and assumptions regarding incidence, he contends that a shift toward income taxes would result in a modest decline in the regressivity of the local tax system.

On the crucial criticism of the property tax as a source of funding for education—that there are significant disparities in tax bases across districts—both McGuire (2001) and Oates and Schwab (2004) present empirical evidence that the income tax base is distributed more evenly across space than is property wealth. McGuire studies the distribution of adjusted gross income per capita and equalized assessed property value per capita across municipalities in Illinois in 1997, and she finds that the coefficient of variation in adjusted gross income per capita is 0.8 compared to a far larger coefficient of variation in equalized assessed property value per capita at 2.8. This much larger variation in the property tax base is attributable in part to the uneven distribution of commercial and industrial property across communities. Oates and Schwab examine income per pupil and property wealth per pupil across school districts in four states (Massachusetts, New Jersey, New York, and Texas) in 1990 and find property wealth per pupil to be one and a half to two times more unevenly distributed across space than income per pupil. Still, because the income tax base also varies across districts, inequities in spending would nevertheless emerge if school districts relied on local income taxes.

Two practical considerations, however, heavily favor the property tax. First, income and sales taxes are highly pro-cyclical revenues, whereas property taxes are more stable over the economic cycle. The latest evidence on the relative stability of the property tax comes from the Great Recession. While state and local government income tax revenues fell by 17 percent from their prerecession peak and sales tax revenues fell by nearly 11 percent, property tax revenues declined by only 2 percent. It is problematic to fund education expenditures, which vary with characteristics of the population and not the business cycle, using economically volatile revenues. Second, the property tax is a highly productive revenue source. For example, in Illinois in 2011 school districts raised $14.7 billion in property taxes. In comparison, the state raised $7.4 billion through the state general sales tax and $14.2 billion through its individual and corporate income taxes combined. If the local property taxes for schools were replaced with a local income or sales tax or with state aid financed through either state sales or income taxes, the required combined sales or income tax rate would be extraordinarily high and could potentially generate large distortions in economic behavior.

What is the Future for the Property Tax in School Funding?

As discussed, there are two fundamental sets of criticisms of the property tax as a source of funds for schools. The first is that because the tax is a local revenue source and the base is unequally distributed across local jurisdictions, it often leads to unacceptably large interdistrict disparities in per-pupil spending. The solution to this problem is to reduce or eliminate reliance on the tax at the local level. The recent shift from equity to adequacy as a focus of court challenges, however, may mitigate the force of this argument. In contrast to policies designed to reduce wealth-related spending disparities, a policy aimed at raising the spending of low-spending school districts to an adequate level need not require a concomitant decline in spending by high-
spending districts. If the dominant goal of a state government’s school finance policy is to ensure that all its students are provided with an adequate education, the state may be willing to allow individual districts to supplement the state-guaranteed adequate level of spending by levying local property taxes, even if the property tax base is unevenly distributed across jurisdictions.

The other set of criticisms pertains to the nature or structure of the tax itself—that it is inefficient, inequitable, and difficult to administer. These criticisms are not specific to the funding of school districts, and in large part they would pertain to the property tax whether it is a local or a state source. Nonetheless, if the property tax comes up short on the criteria of equity, efficiency, and simplicity, the primary reliance of local school districts on the tax is called into question. We think that, with the exception of simplicity and ease of administration, these criticisms are largely without merit. In comparison with local income and sales taxes, the property tax holds its own in terms of efficiency and equity. Of course, the tax in practice is often not a theoretically pure tax. Like all taxes, its ability to score well on economic criteria depends on the base being broadly defined, the marginal tax rates being relatively low, and the tax administration being transparent.

Many property taxes operating today do not meet these requirements. Instead, the base is reduced through preferential treatment of residential property relative to commercial and industrial property; nominal tax rates exceed effective tax rates because assessments are far below market values; and the price of local services is not clear to taxpayers. In addition, access to the tax by local governments is complicated where states have imposed (or voters have supported) limitations and restrictions on the tax. Still, using standard normative criteria to assess its structure and design, we believe the property tax, in practice, holds up well compared to its main alternatives, especially as they would be implemented, and likely compromised, in practice. Although the property tax clearly fails with regard to simplicity and transparency, its productivity as a revenue source and the stability of the tax base over the economic cycle give it two important practical advantages over income and sales taxes.

We conclude that the property tax will continue to have an important role, actually two roles, to play in financing schools. First, the property tax can be used as a source of revenues for state school aid programs either through a statewide property tax or through state capture of locally raised property taxes. Moving from reliance on local property taxes to a state-level property tax has been championed by Sheffrin (2001). He argues that it is hard to justify reliance on the local property tax for funding schools given efforts to “delegitimize differences in educational support across districts” (p. 316). He reasons that a state-level property tax might alleviate concerns of unfairness. Sheffrin notes that a statewide property tax is, in fact, what we have to varying degrees in states as varied as California and New Hampshire. Second, to enhance efficient choice of service levels, the property tax can be a marginal revenue source—to supplement state aid—at the local school district level.

Despite decades of criticism, the property tax remains the major source of local tax revenue for schools. Given that policymakers would find fault with any local tax alternative for funding education, the property tax must play an important, if perhaps diminished, role in funding schools now and into the foreseeable future.

Notes

1. Several states have considered tax and expenditure limits (TEL) in recent years. Since 2006, nearly a dozen states considered adoption of a TEL. The National Conference of State Legislatures tracks TEL actions. Their most recent report is available at http://www.ncsl.org/documents/fiscal/TELLs2012report.pdf.

2. Authors’ calculations using data from Table 1 of Public Elementary-Secondary Education Finances, Annual Survey of Local Government Finances, 2010–11 (U.S. Census Bureau, 2013). Available at http://www.census.gov/govs/school/.

3. Hawaii has a single state-operated school district.

4. Authors’ calculations using data from Table 4 of Public Elementary-Secondary Education Finances, Annual Survey of Local Government Finances, 2010–11 (U.S. Census Bureau, 2013).

5. See the chapter by Downes and Figlio in this volume for a discussion of the impacts of tax and expenditure limits on school funding and education quality.
Papke (2005, 2008) describes the effects of the 1994 reforms on K–12 school funding and student performance. The property tax rate is oft en expressed in term of mills, or dollars per thousand dollars of assessed valuation. See the Michigan Department of the Treasury (2002) for detailed discussion of this reform. These figures come from the authors’ calculations from the state-level census data. In 2011 aggregate property taxes in California generated $53.5 billion compared to the state individual income tax, which generated $50.5 billion. That local reliance on property taxes is more likely to be tolerated by the courts when the goal is adequacy rather than equity finds support in Figlio (2004) and Yinger (2004). Figlio notes, “adequacy definitions are entirely child-centered, rather than taxpayer-focused ...” (p. 90), thus implying that equity among taxpayers is not necessary, and Yinger concludes that, with adequacy as the goal, a (well-designed) foundation plan would appear to be sufficient.

The figures are restricted to the 36 states with independent school districts. Of these 36 states, 24 did not experience court-mandated school reform, five experienced court-mandated school reform before 1985, and seven experienced court-mandated school reform after 1985. Evans, Murray, and Schwab (2001) interpret adequacy along these lines. They contend that adequacy can be achieved through a well-designed foundation aid program coupled with local control over spending above the amount required for an adequate education.

Yinger (2004) discusses the various ways in which states supplement foundation programs to achieve greater equity. Loeb (2001) argues that a system of state grants coupled with unlimited ability to raise additional revenues at the local level may not be politically sustainable.

While the portion of the property tax that applies to land (which is clearly immobile) is arguably nondistortionary, the portion that applies to buildings and other structures is likely to discourage taxpayers from improving their property. See Netzer (1998) for a discussion of the rationale for and issues surrounding land taxation.

See Mieszkowski and Zodrow (1989) and Zodrow (2001) for interpretive surveys of the relevant literature. Dye, McMillen, and Merriman (2006) describe and evaluate a rather byzantine assessment limit on residential properties in Cook County that was championed by the county assessor, an elected official.

Approximately 200,000 taxpayers were subject to the AMT in 1980. By 2005, 5.5 million taxpayers paid the AMT. In late 2012, Congress passed legislation that permanently indexes the AMT for inflation. According to estimates by the Tax Policy Center (available at www.taxpolicycenter.org/numbers/displayatab.cfm?DocID=3967) in 2013, about 3.9 million taxpayers were subject to the AMT, with the number expected to rise to 6.1 million by 2023.

This effect may be much weaker than anticipated. After the Tax Reform Act of 1986 eliminated deductibility of state sales taxes but retained deductibility of state income taxes, many commentators predicted that states would reduce reliance on sales taxes in favor of the deductible income tax. Instead, the vast majority of state tax increases post TRA86 involved increases in sales taxes.

The well-known flypaper effect, whereby recipient governments spend a much greater proportion of a grant dollar than a dollar of local income, is a likely manifestation of this phenomenon. See Hines and Thaler (1995).


McGuire (2001) provides evidence that the sales tax, with a coefficient of variation of 2.6 in sales per capita, is distributed as unequally across space as the property tax.


These percentage revenue reductions were calculated using data from the Census Bureau’s “Quarterly Summary of State and Local Tax Revenue.” Revenue totals are for the 12-month period ending in each quarter. Percentage changes were calculated using data for the same quarter in the relevant years.

Note that this criticism is not per se about the property tax; rather it applies to any local tax whose base is unevenly distributed.

See Evans, Murray and Schwab (2001), McGuire (2001), Sheffrin (2001), and Oates and Schwab (2004) for similar conclusions on the continued importance of the property tax.

Ladd and Hansen (1999) also argue the case for state-level property taxes on equity grounds.
References


taxes, taxes on land rents, and property taxes. *Journal of Economic Literature, 27,* 1098–1146.


23
Tax and Expenditure Limits, School Finance, and School Quality

THOMAS A. DOWNES AND DAVID N. FIGLIO

Introduction

While the preponderance of tax and expenditure limits (TELs) currently in place date back to either the “tax revolt” of the late 1970s and early 1980s or to a second mini-revolt in the early 1990s, the combination of rapidly escalating real estate prices prior to the Great Recession and continued growth in property tax bills during the Great Recession led to numerous recent efforts to control the growth of state and local revenues. Since 2004, Taxpayers’ Bills of Rights (TABORs) have been promoted in 30 states. While none of these efforts have been successful, and the model for these proposals, Colorado’s TABOR, was rolled back in 2005, these efforts make clear that the goal of these TABORs—to provide tax relief for populations feeling increasingly burdened by tax collections, particularly the property tax—retains considerable popular support. Thus, while few new TELs have been enacted in the 21st century, the fiscal environment throughout the United States appears ripe for new fiscal constraint proposals. In fact, New York State enacted a property tax cap in 2011.

As noted, the best known “tax revolt” in the United States occurred in the late 1970s and early 1980s, when one in three states (most notably California, whose Proposition 13 was not the first but was the most influential of the modern tax and expenditure limitations) acted to limit revenue or expenditure levels or growth. Like the recent push for TABORs, the second mini-revolt in the early 1990s was prompted in part by an economic downturn. By the end of that spate of TELs, another ten states either enacted or strengthened their fiscal constraints. By early 2012, 39 states had active tax and expenditure limitations (TELs) in place.

Tax and expenditure limitations are important for education finance and policy because education spending is today the single largest component of state and local budgets. K–12 education spending accounts for almost 22 percent of combined state and local expenditures, and higher-education spending accounts for about 10 percent of state and local spending. If TELs restrict the size or growth of state and local budgets, they may lead either to reductions in education spending or to reallocations in spending from one education category (say, higher education) to another education category (such as K–12). Some state TELs, such as in Oregon, explicitly “protect” state spending on K–12 education at the expense of other state budget categories. This protection of state spending on K–12 education has led to dramatic cutbacks in higher-education spending in some states (Archibald & Feldman, 2006). Still, even though K–12 education spending has been protected relative to other categories of spending, the level of K–12 education spending and the distribution across school districts of this spending may have been influenced by TELs. In the case of local education agencies that are their own taxing authorities, TELs that limit their revenue or spending either reduce spending on education or change the mix of revenues used to finance education spending. TELs can also influence the distribution across school districts of education spending as well, as some school districts are more likely to be constrained by TELs than are others. It is therefore important to understand the degree to which TELs affect overall revenues for, and spending on, K–12 education and the distribution across districts of revenue and spending.

Understanding the impact of TELs on both the average level and distribution of educational outcomes such as student achievement is of equal importance. Even if TELs reduce education spending, they may not reduce achievement. Studies of the 1970s-era limitations (e.g., Citrin, 1979; Courant et al., 1980; Ladd & Wilson, 1982; Stein et al., 1983) indicate that supporters of TELs believed that these limits would lead to increased efficiency in government; that is, that governments would cut waste without reducing service levels. On the other hand,
analysts of TELs have argued that these policies could lead to reductions in student outcomes that are far larger than might be expected given the changes in spending. The literature we summarize below indicates that this latter outcome is most consistent with the data.

In this chapter, we first briefly describe the different types of TELs active in the United States today. We then summarize the literature on the relationship between TELs and the fiscal structure of local governments and provide an overview of the literature that has examined the impact of limits on student performance. After offering potential explanations for why student outcomes might be influenced to a different degree by TELs than by spending changes in general, we close with a discussion of gaps in our knowledge.

**Tax and Expenditure Limitations in the United States**

TELs that affect school districts and other local governments take a variety of forms that differ in their restrictiveness. The most common form of TELs prior to the 1978 passage in California of Proposition 13 was a limit on specific property tax rates. For any local jurisdiction, such as a school district or a water authority, the property tax revenues of that jurisdiction are determined by the following relationship:

\[
\text{Property tax revenue} = (\text{Tax rate levied by the jurisdiction}) \times (\text{Taxable value of property}).
\]

Total tax payments made by the residents of any jurisdiction equal the revenues of that jurisdiction and any payments made to overlying jurisdictions. As a result, limits on specific tax rates, that is, setting a maximum tax rate that specific local jurisdictions can levy, may not reduce the property tax payments of residents of those jurisdictions. This follows, in part, because spending responsibilities could be shifted from the jurisdictions constrained by the specific limit to overlying jurisdictions that are not constrained and in part because the limits do not control the growth in the taxable value of property in the jurisdiction, also known as the assessed value of property.

Other TELs limit the overall property tax rate levied by all overlapping taxing authorities within a geographic area, but, as with specific tax rate limits, they can be circumvented with changes in the assessed value of property in the affected jurisdictions. The majority of post-1978 TELs either limit both property tax rates and growth in the assessed value of property or directly limit the growth in property tax revenues. Less common but potentially more restrictive still are limits on overall revenue growth from all sources, or limits on expenditure growth.

These "modern" TELs are more likely to be effective in restricting the total revenues and expenditures of all jurisdictions within a geographic area than are the earlier forms of limits. The extent to which a limit constrains the ability of local governments to raise revenues or make expenditures depends not just on the type of limit, but also on how restrictive it is. For instance, a TEL that limits annual revenue growth to 1 percent is much more likely to constrain local governments than would a TEL that limits growth to 10 percent. Figlio (1997) and others have attempted to quantify variation in the extent to which limits constrain local governments, and (as discussed later in this chapter) have found that variation in the extent to which limits constrain matters less for student outcomes than the presence of any limit that could be “binding”; that is, that could impose an effective constraint on local fiscal behavior.²

At the time of writing, nine states (Arizona, California, Colorado, Iowa, Kansas, Maine, Minnesota, Nevada, and New Jersey) impose restrictions on aggregate spending by all local governments, and Wisconsin imposes limits on nonfederal revenues of school districts. Other states limit property tax revenues but do not restrict total revenues (i.e., revenues from all sources including intergovernmental aid) of all local governments; these states are Idaho, Indiana, Kentucky, Louisiana, Michigan, Mississippi, Missouri, New Mexico, Ohio, Oregon, Texas, Utah, West Virginia, and portions of Illinois in the Chicago metropolitan area. Several of the states with overall spending limits, such as Colorado and Minnesota, also limit total revenue of all local governments. Arkansas, Florida, Maryland, Oklahoma, and Washington do not directly limit either revenues or expenditures but effectively do the same thing with combinations of tax rate limits and limits on property assessments. It is not likely that the TELs imposed in other states have had any real effect on revenue.
The Effect of TELs on Education Revenues and Spending

The most natural place to begin a discussion of the effects of TELs on schools is to consider whether TELs have influenced school finances at all. McGuire (1999) and Mullins and Wallin (2004) summarize much of the early literature on the fiscal effects of TELs. TELs apparently slowed the growth of property tax revenues, but did not significantly slow the growth of total revenues received by all local governments. Indeed, the most compelling estimates indicate that TELs have little or no effect on total revenues of local governments, where revenues are aggregated across all local governments. The research also points to three changes in the fiscal systems in states with local limitations that have blunted the effects of TELs on the revenues of local governments. First, increased state aid, funded by increased state taxes, compensates for the decline in local revenues (Mullins & Wallin, 2004). Second, overrides, in the states that permit overrides, have permitted voters in some localities to allow post-limitation spending to continue along the path it would have followed in the absence of the limit (McGuire, 1999). Third, in localities constrained by TELs, there has been an increased reliance on other local taxes and user fees (Mullins & Wallin, 2004).

Other research confirms the results of the earlier work and provides a richer picture of the effects of TELs on the revenue and expenditure of local governments. Shadbegian (1998, 1999, and 2003) documents the varied effects of state and local tax and expenditure limitations on revenues and expenditures of local governments. Examining the impact of TELs on all local revenue sources and accounting for differences in limit stringency, he finds that TELs reduce per capita property tax revenue, local own-source revenue (i.e., revenue from all sources other than intergovernmental aid), and local expenditures. TELs reduce local reliance on the property tax and on own-source revenues, though an increase in the use of miscellaneous revenues mitigates the extent to which the decline in property tax revenues translates into a decline in own-source revenues. Downes and Killeen (2014) confirm this finding for school districts, showing that per-pupil revenues from fees and alternative revenues per pupil are higher in school districts subject to TELs. However, school districts appear to make more use of these revenue sources than do other local governments subject to limits.

Shadbegian also finds that the estimated impacts of TELs vary, however, with the stringency of the limits. Localities in states with more stringent limits rely more on other taxes and less on miscellaneous revenues to compensate for the decline in property tax revenues. Shadbegian (2003) narrows the focus to the impact of state and local limits on the provision of public education. Using data aggregated to the state level, Shadbegian establishes that, on net, TELs have almost no effect on total spending on public education because increases in direct and indirect spending by state governments compensate for the decline in spending from local own-source revenue. Not surprisingly, the presence of a limit on the state government revenues or expenditures reduces the extent to which state spending can compensate for reductions in local own-source revenue.

Research also indicates that, while TELs had limited impact on the size of the state and local sectors, the limitations did affect the relative importance of different sources of revenues for local governments. For example, Joyce and Mullins (1991), Mullins and Joyce (1996), and Mullins (2004) find that limits increase centralization of revenue raising, state government spending growth, and growth in the local use of nontax revenues. In addition, TELs reduce the share of revenue raised through local broad-based taxes. The variation across localities in the extent to which the limits constrained local spending (Figlio, 1997) means, however, that state-level data provide an imperfect picture of the full impact of TELs, since spending data from constrained localities are averaged with spending data from localities unaffected by the constraints. Nevertheless, the results from this series of papers establish that, for all local governments, the extent to which TELs affect spending on education or on any other publicly provided service depends upon the ease with which that government can generate nontax revenue and the change in the amount of state aid distributed to that government.

Mullins (2004) uses county-level fiscal data on within-county variation in spending by local governments from all counties in the United States to determine whether TELs affect some parts of a state differently from others. He finds that, on average, TELs increase the within-county variation in both general and education expenditures. The more binding the limit is, the more within-county inequality grows after the TEL. Given that TELs are more likely to bind in initially low-spending districts, TELs may lead to greater disparities in spending across districts within a state because state aid has typically not fully compensated for the impact of the constraints.
Mullins also finds that the impact of TELs on affluent suburban fringe counties differs from the average impact of these limits. For these counties, more binding limits reduce variation in education expenditures across school districts and have no impact on the within-county variation in general expenditures. Why these counties look different from all others is not clear, though differential ability to pass overrides and differential access to nontax revenues are two possible explanations.

Case studies of the evolution of local public spending in California provide some final useful evidence on the impact of TELs. Proposition 13, the voter-approved initiative that was the first in a series of local and state revenue and expenditure limitations imposed on governments in California, limited the overall tax rate on any property to 1 percent of the taxable value of that property. The initiative also limited the growth in the taxable value of property to at most 2 percent, unless that property changed hands, at which time the taxable value would be set equal to the sales price of the property. Although local and state revenue and expenditure limitations coexist in a number of states other than California, the stringency of these key provisions of Proposition 13 as well as the state and local limitations that have since been approved by California voters resulted in a context in California that was unique—at least until the passage in 1992 in Colorado of the Taxpayer’s Bill of Rights. As importantly, the elapsed time since Proposition 13 has made it possible in the California case to examine the long-term effects of stringent fiscal limits. For that reason, a brief summary of the research focusing on California is warranted.

In general, the evidence from California indicates that the evolving fiscal structure of its state and local governments has constrained the ability of cities and counties to increase spending on functions that might substitute for education spending. The evidence from California also sheds light on the extent to which TEL-induced restrictions on education spending are circumvented by increases in noneducation spending by overlying jurisdictions with access to unconstrained revenue sources. In California, a combination of legislation and voter initiatives has resulted both in a reallocation of property tax revenues from cities and counties to school districts and in earmarking portions of state revenues for school districts (Wassmer, 2006). This favored status of school districts, relative to cities and counties, has limited the flexibility of cities and counties and has accentuated the extent to which cities and counties rely on nontraditional revenue sources (Hoene, 2004; Wassmer & Anders, 1999). These trends in the fiscal status of cities and counties have apparently limited the relative growth in expenditures by these governmental units on such services as libraries and parks and recreation, which could substitute for declines in inflation-adjusted education spending. While city and county governments have greater access to nontraditional revenue sources, this access has been used primarily to maintain the share of noneducation spending. Hoene (2004), for example, shows that the shares of total expenditures by California cities on libraries, parks and recreation, and social services have stayed constant or declined slightly since 1977. For California cities, housing and community development was the only spending category that might substitute for education for which the share of total expenditures increased. And for the period from 1990–91 to 1994–95 per capita expenditures by county governments in California were not higher in those counties in which a larger percentage of the population was nonadult (Wassmer & Anders, 1999), even though a larger percentage would have been expected if substitution had occurred. While California’s experience with TELs is unique, the weight of the evidence from elsewhere indicates that California mirrors the general national experience with TELs.

**TELs and Student Performance**

If TELs reduce both the revenues that school districts and other local government generate and the amount that these governments can spend, it is not obvious that these limits would adversely affect student outcomes. As mentioned in the introduction, one primary rationale for TELs cited by supporters is that they could lead to more efficient provision of services. In the case of education, this logic might imply that resources may be deployed in times of fiscal constraint to preserve instructional spending rather than administrative spending. Evidence regarding general school reactions to fiscal stress (e.g., Ladd, 1998) suggests that fiscally constrained districts are more likely than other districts to cut back on administrative and capital spending in order to preserve instructional spending. While this result need not imply that the districts are becoming more efficient (i.e., are able to produce improved student outcomes with the same level of spending), the result does mean
that it is theoretically an open question as to whether TELs would lead to reduced school service levels or student performance.

Even if inefficiency is not pervasive, if the argument made by Hanushek (1986, 1997, and 2003) that increases in spending per se are unlikely to result in large improvements in student performance is correct, then cutbacks in spending could well generate no adverse effects on student outcomes.

An obvious alternative is that the cut resources are productive, so resource cuts would result in reductions in student outcomes. Further, if the resource cuts were concentrated in the most productive areas of education spending, the declines in student performance might be larger than would be expected given the magnitude of the declines in education spending. As a result, assessing the impact of TELs on student performance can only be done by estimating directly this impact.

Early case studies of TELs like those of the Joint Budget Committee (1979) and Schwadron (1984) for California and Greiner and Peterson (1986) for Massachusetts consistently show that residents of the studied states perceived a drop in the quality of publicly provided services. That this perception reflected reality is sometimes, though not always (Greiner & Peterson, 1986), confirmed by objective measures of service quality. According to the case studies, government officials responded to TELs by first making cuts in capital expenditures and in areas of current expenditure that these officials felt were peripheral. For example, in California school administrators sought to protect the core academic subjects, choosing instead to make cuts by pursuing such strategies as reducing the diversity of course offerings and the number of pupil service employees.\(^5\)

Given their timing, these case studies could not be used to draw any conclusions about the long-run effects of TELs. Also, even though these case studies moved beyond examination of the fiscal impacts of TELs, the absence of any clear prediction about the impact of TELs on student performance meant that only by examining student outcomes directly, and by determining how these outcomes had changed relative to the pre-limit baseline, could researchers ascertain the effect of limits.

Downes (1992) offers the first over-time comparison of the effects of TELs on student performance. Comparing performance on the California Assessment Program test for 170 unified (K–12) districts in 1976–77 and 1985–86, Downes finds that student performance actually increased by five points, on average, and that the cross-district distribution of student performance was essentially unchanged over this time period. This research implies that California’s Proposition 13 did not reduce student performance at any point on the performance distribution.

Such a conclusion is unwarranted, however. Contemporaneous with the state and local response to Proposition 13, the state implemented school finance reforms made necessary by the Serrano v. Priest decision of the California Supreme Court.\(^6\) This reality raises a problem that faces any researcher attempting to isolate the impact of TELs on public schooling. Frequently, states have implemented major school finance reforms close in time to the passage of TELs. Thus, the effects of TELs can be isolated only by looking across the states or by examining the long-run experience in a state in which a limit was passed and no major changes in the school finance system had occurred (Blankenau & Skidmore, 2004).

Two papers take this lesson to heart and, thus, provide a model for future empirical research on the impact of TELs. Using a cross-section of student-level data from the National Education Longitudinal Survey (NELS), Figlio (1997) finds that, ceteris paribus, revenue and expenditure limits were associated with lower 10th-grade performance in mathematics, reading, science, and social studies. Using variation generated by the imposition of property tax limits on some, but not all, school districts in Illinois, Downes, Dye, and McGuire (1998) conclude that, in the short term, these limits led to slower growth in the performance of third graders on a standardized test of mathematics, but did not do so for eighth graders or in either grade in reading.

While these papers provided improved estimates of the effects of TELs, they had potential flaws. For example, while he introduced a new instrumental variables strategy to predict which states would implement TELs and ruled out many competing alternative explanations for his findings, Figlio’s (1997) work drew upon cross-sectional comparisons, making it impossible for him to fully rule out the possibility that the school districts subject to TELs were a nonrandom sample of all U.S. school districts.\(^2\) Downes, Dye, and McGuire’s (1998) comparison of cross-time changes in performance in districts subject to TELs to similar changes in Illinois districts not subject to TELs could have resulted in flawed estimates of the impact of TELs because of the existence of only three post-limit years\(^3\) of data and the possibility that the limited districts might be a
nonrandom selection of Illinois districts. Nevertheless, these two papers clarified that evaluating the effects of TELs requires not only before-and-after data on students in districts subject to limits but also a control group of students from states in which no limits have been enacted.

With this observation in mind, Downes and Figlio (2000) attempted to determine how the TELs of the late 1970s and early 1980s affected the distribution of student performance in states in which limits were imposed and how student performance has changed in these states relative to student performance in states in which no limits were imposed. The core data used in the analysis were drawn from two national data sets: the National Longitudinal Study of the High School Class of 1972 (NLS-72) and the 1992 (senior year) wave of the NELS. The NELS data were collected sufficiently far from the passage of most TELs to permit quantification of the long-run effects of these limits by analyzing changes in the distributions of student performance between the NLS-72 cross section and the NELS cross section. Focusing on two rounds of national datasets allowed Downes and Figlio to implement a differences-in-differences strategy in which they compare performance of students in states after versus before they implemented TELs. They further attempted to account for the potential endogeneity of TEL adoption by following the logic of Rueben (1997) and using state constitutional language to instrument for adoption. The estimates thus come as close as possible to quantifying the long-run causal impact of TELs.

Downes and Figlio confirm Figlio’s (1997) finding that TELs reduce mathematics test scores by 1 to 7 percent, depending on model specification, as well as Downes, Dye, and McGuire’s (1998) finding of no observable impact of TELs on reading scores. Given the age of the test-takers, it is sensible to believe that high school mathematics differences may be more attributable to differences in schooling than are high school reading differences. Thus the generally stronger effect of TELs on mathematics than on reading should come as no surprise. Moreover, this pattern of stronger effects for math is fully consistent with Figlio and Ladd’s findings with respect to school accountability programs elsewhere in this handbook.

Downes, Dye, and McGuire (1998) and Downes and Figlio (2000) both find that TELs have more negative effects on student performance in economically disadvantaged localities, though these estimated differences are frequently imprecise and so should be taken with a grain of salt. That said, the general pattern of these results is consistent with the finding that TELs affect low-income communities more than they affect higher-income communities (Bradbury & Zhao, 2009) and the increased likelihood that they would bind poorer communities may exacerbate existing inequalities across school districts.

For the most part, when researchers examine the impact of TELs on student performance, they limit their analysis to students who remain in the public schools. Downes and Figlio (1999a) provide the only attempt to study the performance effects of TELs (and school finance reforms) on private school students. Using a methodology similar to Downes and Figlio’s (2000) study of public school performance, they find evidence of a modest, though imprecisely estimated, negative effect of TELs on student test scores in the private sector. Determining whether these negative effects are attributable to lower, private sector quality because of diminished public sector competition or to changes in the composition of students served by private schools requires analysis of the enrollment effects of TELs. To date, the only such analysis is provided by Bradbury, Case, and Mayer (1998), who analyze the relationship between grade-level enrollment patterns and various indicators of the extent to which TELs affect the ability of affected towns to increase education spending. Since differences between actual enrollment patterns and the patterns of enrollment implied by the decennial censuses reflect primarily withdrawal from the public schools, either to private schools or nonenrollment status, the results from this paper shed some light on the effect of TELs on dropout rates. Bradbury, Case, and Mayer find that the share of the potential student population served by the public schools is lower in districts in which more initial cuts were necessary when the limits were first imposed. This result suggests that limits could increase dropout rates or could result in students switching from the public to the private sector. Further research on the impact of limits on school completion and on the share of students attending private schools is clearly needed.

If Money Does Not Matter Much, Why Might TELs Result in Quantitatively Large Performance Reductions?
The weight of the evidence on TELs suggests that these limits lead to small reductions in school spending and quantitatively large (and statistically significant) reductions in student performance. Since even those studies that appear to establish that "dollars matter" indicate that the marginal value of a dollar is small, this seemingly asymmetric result, small declines in spending appear to translate into relatively large declines in performance, is surprising.

Interestingly, most of the explanations in the literature for little relationship between adding spending and increased achievement do not have obvious symmetric implications for student outcomes when spending is reduced. For instance, results suggesting no relationship between additional spending and student outcomes are consistent with additional dollars being captured as union rents or administrative rents, with no additional productive inputs being provided. By "rents," we mean that individuals currently in control of resources (either administrators, or experienced teachers, through bargaining in the union) may be able to direct new resources to expenditure categories that most benefit them personally, rather than choosing the most effective use of additional resources. If rent-seeking is an explanation for the small gains associated with increases in expenditures, relatively large reductions in outcomes could result when equal-sized reductions in resources are imposed, as is likely in the case of TELs. Consider, for instance, the case of administrative rents. Reductions in spending would be more likely to be borne by instruction, rather than administration, just as increases in spending accrued to administration, rather than instruction. If instructional spending matters for student achievement per se, then this logic would imply that reduced spending, leading to reduced instructional spending, would have impacts on student outcomes that are larger than would similarly sized spending increases. Note, also, that reductions in spending that lead to cuts in instruction, rather than administration, are also consistent with a story of an efficiently run school already operating with the minimal necessary fixed administrative costs. In such a situation, even an efficiently run school would need to cut its instructional budgets, which would, in the current example, harm student outcomes.

In general, if reductions in spending do not change (or exacerbate) the premium to a less productive input, then outcomes should fall, just as outcomes would not be expected to rise as spending increases. This observation implies that the case of teacher union rents is similar to the administrative rents case. If reductions in spending increase the relative salary of experienced teachers in a district because starting salaries are cut, the school district may end up holding on to its more experienced teachers while running off its less experienced teachers and reducing the ability to attract high-quality teachers from the outside.

If teacher union rents are a partial explanation of the findings of a negative relationship between tax limitations and student outcomes, then one should expect that tax limitations are associated with changes in the teacher labor market. Figlio and Rueben (2001) find that tax limitations substantially reduce the average qualifications of new entrants into a state’s teaching force. Several researchers, such as Ehrenberg and Brewer (1994) and Goldhaber and Brewer (1997), demonstrate that teacher characteristics such as the selectivity of the teacher’s undergraduate institution, as a proxy for native ability, or teacher subject matter expertise have statistically significant influences on student outcomes. Thus the finding that these teacher attributes fall with tax limitations is consistent with a reduction in student achievement. This evidence could also partially explain the differences between Downes, Dye, and McGuire’s (1998) results suggesting relatively small achievement effects of TELs and Downes and Figlio’s (2000) results suggesting larger achievement effects because the teacher labor market responses to TELs should be relatively long-term.

Moreover, Figlio and Ruenen’s (2001) results are consistent with an explanation of union rents, because they find that teacher quality responds to TELs by even more than what would have been predicted by changes in teacher salaries alone. Further evaluation of different data, though not part of the Figlio and Ruenen paper, suggests that school districts subject to TELs tend to cut the salaries of starting teachers while maintaining teacher salaries for more experienced teachers. This finding helps to explain why TELs apparently led to reduced teacher quality among new teachers that is larger even as the salary reductions per se associated with tax limitations were comparatively modest.

Asymmetry in the impact of TELs could also result because school districts might not respond to budget reductions in a manner consistent with increasing efficiency. Even if schools in the districts affected by the limits are capable of reducing waste and thus maintaining the pre-limitation level of student performance, the limits typically provide no explicit incentive to administrators of these districts to eliminate waste. Figlio and O’Sullivan (2001) provide evidence of strategic behavior of municipalities, townships, and school districts in
response to the imposition of a limit. One goal of this behavior seems to be to encourage voters to override the limit. For school districts, if school district administrators are budget maximizers and if technical inefficiency in the schools persists because insiders in school systems have more information than outsiders about how resources can be used productively, then administrators may have an incentive to allow student performance to decline by more than is dictated by changes in financial circumstances and to use this decline to argue for additional resources.

Figlio (1998) finds results consistent with this argument, showing that in school districts bound by TELs, measured school inputs are reduced but administrative expenditures are not. Similarly, Figlio (1998) establishes that school districts affected by Oregon’s Measure 5 are more likely to cut instructional expenditures than administrative expenses.11

Dye and McGuire (1997), on the other hand, exploring the short-term effects of a tax limitation affecting only portions of Illinois, find very little evidence to suggest that school districts cut instruction to preserve administration expenses; if anything, affected school districts may have cut administration more. This finding is wholly consistent with Downes, Dye, and McGuire’s (1998) evidence that this Illinois tax limitation has not appreciably affected student outcomes in the constrained school districts, at least in the short run. It is possible that the administrative-rents explanation can provide some reconciliation between the findings of little achievement implications in Illinois versus larger achievement implications found using other methods from a national study. One factor that might explain why Illinois school districts did not cut instruction, relative to administration, following their tax limitation, while other school districts facing other limitations apparently did, involves Tiebout (1956) competition. The attributes of the Illinois tax cut that facilitate nice quasi-experiments—that is, the fact that some Chicago-area school districts were affected while others were not—also makes it possible that the unaffected school districts could put competitive pressure on the affected districts. This finding is confirmed by Millimet and Rangaprasad (2007), who find evidence of competition among Illinois school districts only in those counties in which school districts are subject to TELs.

While Dye and McGuire (1997) indicate that, in instances in which interjurisdictional competition might provide incentives for administrators to use resources efficiently, instructional spending might be protected, Dye, McGuire, and McMillen’s (2005) update suggests that the discipline provided by competition may disappear when the limits become more widespread. Over time in Illinois, voters in many more school districts have chosen to adopt limits. Using data on both the early and later adopters, Dye, McGuire, and McMillen find that, in the long run, instructional spending is not protected. Such a result could indicate that the more widespread adoption of limitations in the Illinois context has eliminated some of the competitive discipline that existed when limits applied to a small subset of districts in the state. Thus, in the aftermath of limits, competition may constrain rent-seekers only if the limits do not apply universally.

The magnitude of private responses to constraints provides additional indirect evidence on the limited scope of competitive pressure as a disciplining mechanism. The central lesson from the work on private schooling and private contributions seems to be that the private responses to constraints tend to be small. For example, Brunner and Sonstelie (2003) note that, while private contributions are a substantial portion of district revenues in a small number of school districts, overall these contributions have little impact on the post-Serrano distribution of education expenditures. Brunner and Imazeki (2003) confirm this finding based on data from 2001. They show that less than one-fifth of California students attend schools where combined contributions to the school and district exceed $50 per pupil. In Vermont, a court-mandated finance reform in 1997 imposed fiscal constraints on some but not all school districts. Downes and Steinman (2006) took advantage of this variation to show that while contributions may have been slightly muted, the impact of constraints and the overall effect of contributions on the distribution of spending in Vermont was limited. And Nelson and Gazley (2014), using national-level data on school districts, find little growth in private contributions in response to the fiscal constraints created by the Great Recession.

In a similar vein, Anderson and Butcher (2006) demonstrate that fiscal constraints imposed by TELs lead school districts to enter into more junk food contracts in an apparent attempt to boost revenues. They find, however, that the additional revenues have not resulted in improved student performance, while the increased availability of junk food seems to have led to increased obesity.

Aaronson (1999) puts forth an additional explanation for why tax limitations may reduce student outcomes. Aaronson demonstrates that communities constrained by TELs become more heterogeneous over time. If
educating a more heterogeneous population is more costly, then one might expect that fiscally constrained school districts would experience reduced outcomes for any given level of spending. In addition, if reduced spending leads some students to leave the public sector for the private sector (Downes & Schoeman, 1998; Husted & Kenny, 2002; Evans, Murray, & Schwab, 1999; Bradbury, Case, & Mayer, 1998), especially the highest-ability students (Epple, Figlio, & Romano, 2004), mean performance levels in the public sector would fall and the remaining students in the public sector would be costlier still to educate.

Downes and Figlio (1999a, 1999b) present complementary evidence on the school choice responses to TELs and school finance reforms. Using school-district-level data, they find that school finance reforms, and to a smaller extent TELs, reduce the share of students from high-income families attending public schools and increase the share of such students attending private schools. This result, coupled with those of Aaronson on community composition and of Lankford and Wyckoff (2000) and Figlio and Stone (2001), which suggest that factors correlated with TELs, such as student-teacher ratios, are important determinants in private school selection for relatively high-socioeconomic status families, indicates that tax limitations have the potential to dramatically alter the student composition in the public schools. This resulting change in school composition can help to explain why tax limitations might affect student outcomes (for the students remaining in the public schools) to a greater degree than would be predicted from the corresponding reduction in the level of spending.

In the space available, we cannot offer a comprehensive explanation of all of the possible causes of the negative relationship between TELs and student outcomes. Nevertheless, the preceding discussion demonstrates that it is plausible that student outcomes could fall with the reductions in spending induced by TELs in more dramatic fashion than outcomes would be expected to rise if school districts faced equal-sized spending increases. At a minimum, the results imply that a central desire of voters for TELs—that they would receive reductions in taxes but not face appreciable reductions in student outcomes—appears to have not been realized. Since TELs provide no explicit requirements to spend money in a particular way, there is no reason to believe that an additional constraint would necessarily eliminate waste and inefficiency even if the system was inefficient or wasteful at the outset. Moreover, if school administrators or teacher unions extract rent from the system, tax limitation measures may exacerbate existing inefficiency if the most flexible components of spending are the ones that, if reduced, would result in the largest relative declines in student performance. These results may or may not imply that governments fail to allocate resources to those inputs most strongly linked to student performance. But even if waste exists, the evidence does not suggest that tax limitation measures reduce this waste, or at least reduce waste at low cost.

Discussion

The last 25 years have provided researchers with an opportunity to determine if TELs are good public policy. The proponents of the argument that government insiders seek to maximize the size of government contend that constitutional constraints such as Proposition 13 could reduce the size of local governments without affecting the quality of public services provided. This “free lunch” would be realized, the argument goes, because the limits would force governments to eliminate waste.

In this chapter, we summarize the research that has shed light on the validity of the argument that TELs provide a costless mechanism for constraining government spending. The research we summarize produces a relatively consistent conclusion; the imposition of TELs results in long-run reductions in the performance of public school students. We also provide evidence that reconciles these reductions with the most often reported result of the education production literature: dollars matter little per se, if at all. Still, whether TELs are good public policy remains an open question, we contend, since work remains on the quantification of the costs and benefits of TELs. This quantification requires information on the magnitude of the reductions in tax liabilities that result, on the changes in the mix of revenue sources used, on the impact of limits on student performance and on schooling services more broadly, and on the impact of limits on the distribution of spending across levels of education and nonschooling services. Further, how TELs affect the net well-being of individuals at different locations on the socioeconomic spectrum must be assessed.

As this discussion reveals, we know the most about the impact of TELs on tax liabilities, on the mix of revenue sources used, and on the mix of expenditures. Yet, even in these areas, gaps in our knowledge remain.
Assessing the distributional implications of limits requires further analysis, adding to that of Shadbegian (1999, 2003) and Mullins (2004), of the extent to which TELs affect some communities and school districts more than others. Additional work on the impact of limits on nontraditional revenue sources, such as private contributions to fund education and other publicly provided services, is needed. Larger gaps remain in our knowledge of the impact of limits on student performance. Since those gaps have existed, in part, because of the absence of pre- and post-limit data on student performance, the now universal existence of accountability systems that require annual testing of all students should make it possible to assess the impact of newer limits.

Finally, Reschovsky (2004) suggests that the existence of fiscal constraints could serve to exacerbate the impact of downturns on education spending, both by limiting the ability of localities to respond to state aid cuts and by shifting local revenue away from a stable source, the property tax, to less stable sources. Further research into the extent to which TELs make education spending more cyclical is warranted, as is research into the degree to which limits add to the uncertainty of those individuals with the responsibility for crafting budgets.

Notes

1. Anderson and Butcher (2006) review several other potential explanations for voter support for statewide limitations on the ability of local governments to raise revenue or make expenditures.

2. The convention in the literature has been to classify a limit as binding if the limit constrains overall revenue growth from all sources, constraints overall expenditure growth, or combines a limit on the growth of the assessed property value with a limit on the overall property tax rate.

3. Using the terminology of Figlio (1997) and others, all of the TELs considered in Shadbegian (1999) are potentially binding. Shadbegian classifies limit stringency based on the maximum permitted growth rate in nominal property tax revenues.

4. Growth of nontax revenue may be an explanation for Dye, McGuire, and McMillan’s (2005) findings that in Illinois school districts subject to a tax cap, growth in operating spending had declined less than had growth in school district tax revenues and that the different changes in growth rates were not attributable to compensating changes in state aid.

5. The results in Downes (1996) suggest that school administrators in California did not respond to the limits by cutting the administrative staff. For a national cross section, Figlio (1997) also finds no evidence of cuts in administration.

6. Fischel (1989, 1996) makes a strong case that the prospective school finance reforms that were compelled by the Serrano decision stimulated enough additional support for tax limits to make passage of Proposition 13 inevitable. If this logic is right, any observed changes in the distribution of student performance in California should ultimately be attributed to the finance reforms, not the resultant tax limits.

7. Shadbegian and Jones (2005) also use cross-sectional data in their analysis of the impact on student achievement of the stringency of the constraints imposed by Proposition 2½ in Massachusetts. Again, as is the case with Figlio (1997), the possibility exists that some combination of sorting and unobserved tastes for education resulted both in towns being constrained by the limits and having lower levels of student performance. In this particular case, the absence of data on pre–Proposition 2½ test performance means that the authors are unable to rule out the possibility that there existed unobservable factors that resulted in lower test performance and that are correlated with the extent to which a locality was constrained by Proposition 2½.

8. Dye, McGuire, and McMillan’s (2005) finding that the Illinois limits have become more binding over time suggests strongly that the long-term effect of the limits may be different.

9. Downes and Figlio find stronger evidence of a negative effect of TELs on reading scores when they treat TELs as endogenous—that is, that factors influencing reading scores might also influence the likelihood of TELs in a state. See their paper or Figlio (1997) for more of a discussion of the potential endogeneity biases, as well as a detailed treatment of the issue of reverse causality.

10. Epple and Romano (1998) theoretically describe stratification patterns between the public and private sectors that predict that reduced public sector spending leads to the movement of “top” public school students into the private sector, reducing the average performance level of both the private and public sectors. Epple, Figlio, and Romano (2004) offer some empirical justification of the stratification patterns identified in the theoretical model.

11. Shadbegian’s (2003) finding that student-teacher ratios appeared to increase a small amount in states in which local school districts were subject to TELs even though total spending on public education was unaffected provides additional evidence of instructional resources being
reduced in states with binding TELs even when these reductions appeared to be unnecessary.

12. This finding, again, is fully consistent with the stratification patterns predicted in a slightly different context by Epple and Romano (1998), in which relatively high-ability students would leave the public sector for the private sector as the public sector looks less attractive to them, thereby reducing the average socioeconomic status in both the public and private sectors.

References


Section V
Educational Markets and Decentralization
Issues in Educational Privatization

HENRY M. LEVIN

Introduction

Throughout most of the world, private schooling preceded public schools. Private schools were largely religious in their founding and operations and emphasized values and character as much as literacy and secular knowledge. It was not until the middle or latter part of the 19th century that industrializing countries established broad-based government sponsorship and operation of public schools—eventually dwarfing the importance of the private sector in the education of the overall population. In the United States, private or independent schools account for only about 8 percent of enrollments today (National Center for Education Statistics, n.d.a) and more than 80 percent of their enrollments are in religiously affiliated institutions (National Center for Education Statistics, n.d.b).

Towards the end of the 20th century, however, a movement that is broadly referred to as “educational privatization” arose in the United States and in other industrialized or industrializing societies (Levin, 2001). This phenomenon has coincided with a broader and more general, even global, shift of political and economic institutions towards greater reliance on markets in the last two decades. Thus, educational privatization represents just one component of a much larger trend away from government provision of services.

In education, the privatization movement has taken a variety of institutional forms that are discussed in subsequent chapters in this volume. For example, a movement has developed in the nation to establish charter schools that are publicly authorized and financed, but governed by their own private boards. A significant portion of charter schools are managed by for-profit private educational management organizations (EMOs) and nonprofit charter management organizations (CMOs). Public school systems have also joined in contracts with CMOs and EMOs to operate particular district schools independently of conventional district governance and management.

Another form of privatization is embodied in educational vouchers, whereby public funds are provided to families in the form of a certificate or voucher that can be applied towards the costs of private school. While these programs have been slow to take hold in the United States, 13 voucher programs are currently operating (Zimmer & Bettinger, this volume). In some states, taxpayers have been granted tax credits or deductions from taxable income covering some portion of their tuition payments to private schools (Huerta & d’Entremont, 2007; Welner, 2008). The purest form of privatization has been the practice of home schooling where parents reject formal schooling and undertake responsibility for the education of their own children, a growing trend (Cooper, 2005).

Educational privatization is controversial and politically contentious, largely because it reaches into deep ideological and societal divisions in our society (Belfield & Levin, 2005a). The traditional rationale for government funding and provision of schooling is premised on the need to provide a common and equitable educational experience that will prepare all students for their economic, political, and social roles in a democracy. In contrast, privatization focuses on market provision of education, allowing families to choose the type of school that best meets the needs of their own children. In this respect, privatization is mainly devoted to meeting the private goals of families rather than the broad public goals for social and democratic participation in society. As I discuss below, conflicts arise between addressing the private goals of education through choice and the public ones through a common educational experience.

This chapter provides a framework for understanding this conflict between public and private goals and for evaluating different forms of educational privatization. Since this section of the Handbook offers detailed
chapters on educational vouchers (Zimmer & Bettinger), charter schools (Bifulco & Bulkley), and supply of schooling options (Cowen & Toma), the reader is referred to these chapters for details on specific approaches to privatization. Instead, I explore here the commonalities among these phenomena and develop a framework that encourages comparison and analysis of their provisions and consequences. The main themes are that privatization can take many forms and that even within a specific form of privatization, differences in design features can induce vastly different consequences. Finally, the general term “educational privatization” carries little meaning without detailed information on the forms it takes.

Defining Educational Privatization

Defining the term educational privatization is less straightforward than it might appear because almost all forms of education have both public and private components. The term private refers to the provision of schooling for and by individuals, groups, institutions, or entities that is primarily devoted to meeting the private goals of both the school clientele and the institutional sponsors. In contrast, the term public characterizes entities and purposes that are dedicated to a broader societal impact beyond that conferred upon the direct participants, and is usually (but not always) associated with a government role. Given these working definitions of “public” and “private,” there is considerable overlap between them in their application to the educational arena.

Public and private aspects of education can be viewed through at least five key lenses: sponsorship, governance, funding, production, and outcomes.

Educational sponsorship defines who establishes schools and who is ultimately responsible for the schooling that is provided. This is the most common concept used in distinguishing public from private schools. Clearly private schools are sponsored by private entities, but so is home schooling. Public schools, in contrast, are sponsored by government. Charter schools cannot be so easily categorized because they are generally initiated by private groups, but subject to official approval and funding by governments. However, in some cases they are being promoted by large city school districts as a strategy for sponsoring a portfolio of different schools including traditional public schools, charter schools, and those contracted with private organizations, both for-profit and not-for-profit in a portfolio model of schools (Bulkley, Henig, & Levin, 2010).

Governance encompasses the overall authority and responsibility for school operations. In the United States, even private schools and home schooling, although privately sponsored, are jointly governed by both public and private entities. Under the decision of the U.S. Supreme Court in Pierce v. Society of Sisters (1925), state governments are responsible for the regulation, inspection, supervision, and examination of all schools, their teachers, and pupils. Home schooling and private schools, though privately sponsored, are regulated (often lightly) by state governments. Charter schools, though sponsored by privately constituted governance boards and given considerable autonomy from many public laws and regulations, are still subject to the overall authority of their states. Public schools are governed by public authorities under state constitutions and legislatures and by state authorized agencies. The management of schools can be delineated from governance in the sense that even some public and charter schools, though subject to public governance, employ private, Educational Management Organizations (EMOs) and Charter Management Organizations (Cowen & Toma, this volume) to manage their organizations.

Sources of funding for all types of schools can be public or private. Although, in theory, the funding streams for the two forms of schooling types are distinct, in practice they frequently comingle. Home schooling and private schools are funded mainly through the private resources of families. Nonetheless, they may also receive considerable public support. Home schools can receive public funding when attending public schools part-time or enrolling in virtual charter schools where instruction is provided through the Internet (Huerta, Gonzalez, & d’Entremont, 2006). As nonprofit entities, almost all private schools receive public services without paying taxes, and many receive federal funding for the instruction of handicapped, low-income, and bilingual students. States such as Ohio also provide transportation and textbooks to private schools for core academic subjects. When states adopt voucher or tuition tax credit programs, private schools are able to benefit from public funds or tax reductions. Charter schools receive both public funds and extensive private philanthropy as highlighted in Bifulco and Bulkley (this volume) and Baker, Libby, and Wiley (2012). Although the
predominant share of public school funding is from government sources, private funds are also sought through both philanthropy and parental charges for specified services.

With respect to the production of education, schools rely on inputs of students, teachers, administrators, facilities, supplies, and other resources. These resources are provided by both public and private suppliers. The most important input to the educational process, however, is the socioeconomic status of the student’s family. Because the family is the most privatized entity in society, the co-production of education between schools and families must depend upon a major private component, regardless of the sponsorship, governance, or funding of the school organization. Not only are the educational attainments of individual children conditioned heavily by their private circumstances, but the productivity and character of the school as a whole is also influenced by the socioeconomic composition of its students. Thus, the private character of the student body influences the public character of the educational process (Zimmer & Toma, 2000).

Schools also enlist other resources from the private sector to produce education. Public schools purchase textbooks, equipment, maintenance, food services, staff development, and consulting services from private vendors. Even more encompassing is the private role played by EMOs and CMOs (Cowen & Toma, this volume), organizations that have been contracted to operate both traditional public schools and charter schools. Such organizations, usually for-profit, not only manage entire schools, but also provide publicly funded, supplementary services such as tutoring, after-school, and summer-school services.

It is also possible to categorize the outcomes of education into both private and public benefits. Private outcomes benefit primarily the individuals being educated and their immediate families, whereas public outcomes benefit the larger society. Clearly, what students learn in schools confers private benefits in terms of skills, knowledge, values, behaviors, and understanding. These, in turn, enhance their capabilities in the workplace and other settings, increase access to further educational opportunities, and translate into better jobs, earnings, health, and personal satisfaction.

Beyond these private benefits, children’s experiences at school can also provide benefits for the larger society—the main justification for public funding. Under ideal conditions, schools provide students with a common set of values and knowledge to contribute to their functioning in a democratic society. Schools can also contribute to equality of social, economic, and political opportunities among persons of different racial and social class origins—thus making for a fairer society. Schooling is expected to contribute to economic growth and high employment as well as cultural and scientific progress and the defense of the nation.

The debate over the various forms of privatization is driven largely by controversy over how much each type of school contributes to public or private outcomes. Presumably, home schools and private schools—with their dedication to the individual quests of families—are weighted more heavily in the direction of the private benefits. Charter schools in many states and traditional public schools are usually required to ensure greater public outcomes by virtue of their conformance with standardized state curriculum and testing requirements.

Advocates of vouchers and tuition tax credits argue that all types of schools—whether public or private—deliver both public and private benefits, and they point to public schools in wealthy neighborhoods attending to the private goals of their families as much as private schools with similar families. Opponents of vouchers and tax credits, however, suggest that a singular stress on market competition by schools on the basis of their narrow appeals to individual families tends to preclude broader consideration of the public benefits of education. They argue that yielding primacy to the self-serving instincts that drive school choice undermines the democratic functions of schooling that are necessary to build a socially cohesive and more equitable society. But advocates of privatization reply that the political conflicts over the roles and operations of existing public schools blunt their efficiency in meeting the specific and diverse educational needs of their constituents. They conclude that market choice over privatized alternatives would serve children and families more effectively (Chubb & Moe, 1990). These underlying contentions about the virtues of government versus markets drive much of the controversy over recent initiatives for expansion of public funding for private schools and other forms of privatization (Belfield & Levin, 2005a).

**Issues in Educational Privatization**

Central to an understanding of educational privatization is the recognition that much of the controversy
surrounding privatization is rooted in different values and objectives of the protagonists on both sides of the
debate. A scrutiny of the literature and the debates suggest that four criteria for evaluating particular forms of
educational organization are central to the debate over privatization: freedom of choice, productive efficiency,
equity, and social cohesion (Levin, 2002).

**Freedom of Choice**

In a democratic society dominated by free markets, most families value the freedom to choose the experiences
that affect their wellbeing. Indeed, this freedom is viewed as a basic right with respect to childrearing: the
liberty to provide children with the education that is consistent with parental values and educational goals.
Those who push for greater privatization of sponsorship, governance, and public funding for private schools
tend to place great weight on freedom of choice whereby parents select schools for their children and bear the
consequences of those decisions.

**Productive Efficiency**

Education is a costly undertaking, so there is deep concern about how to use resources most efficiently. In
general, there is widespread agreement that policymakers should seek forms of educational organization that
yield the largest educational result for any given resource investment. Proponents of educational privatization
assume that market competition for students among schools will create strong incentives, not only to meet
student needs, but also to improve productivity. Opponents are skeptical of this claim because they believe that
privatized approaches to education will create competition only for the "best" students. Students with greater
needs will have fewer choices and will be isolated in segregated school environments with other students like
themselves. Further, they believe that a preponderance of resources will be used to produce private outcomes at
the expense of publicly valued ones.

**Equity**

A common public goal of education is the quest for fairness in access to educational opportunities, resources,
and outcomes by gender, social class, race, language origins, disability, and the geographical location of
students. Advocates of greater privatization argue that the ability to choose schools will open up possibilities
for students who are locked into inferior neighborhood schools and that the competitive marketplace will
provide greater incentives than those in the existing public school system to meet student needs. Opponents
argue that a move toward greater privatization, such as vouchers, will generate greater inequities because
parents with education and income are better informed and have greater resources for making use of choice,
such as access to transportation. They also believe that the choices themselves will further segregate and disen-
franchise the poor because those with power and status will select schools with student demographics like their
own, and schools will also select students by such criteria.

**Social Cohesion**

The main argument for public funding and operation of schools has been its fostering of a relatively, common,
educational experience that will orient all students to grow to adulthood as full participants in the social,
political, and economic institutions of a given society. Government funding and sponsorship of schools enables
the implementation of a curriculum and pedagogy devoted to a common knowledge base, social values, goals,
language, and political and economic institutions for promoting social cohesion of both youth and adults.
Opponents of greater privatization believe that educational privatization will undermine social cohesion by
giving schools the incentives to differentiate their approaches to enhance their narrow market appeal at the
expense of their social mission (Wolfe, 2003).
The One Best System

The basic policy problem in designing an educational approach comprising these four goals is that they may not be fully compatible with each other. Although it may be possible to establish a plan employing both public and private components, any particular plan will not be able to maximize results on some of the criteria without sacrificing performance on others. In this sense the “one best system” is that which conforms most closely to the preferences of individuals or social groupings regarding the four dimensions of choice, efficiency, equity, and social cohesion. Unfortunately, different groups in different settings will make widely varying judgments regarding the relative importance of each dimension and how each would fare under different forms of public/private organization (Belfield & Levin, 2005a). This complexity can be illustrated by considering the specific designs proposed for publicly funded, educational vouchers. Similar criteria can be used for consideration of the design of charter school plans. Three prominent design tools are: (1) finance, (2) regulation, and (3) support services.

Finance. Finance refers to the monetary value of the educational voucher or public funding of charter schools, how it is allocated, and whether families are able to supplement the voucher to send their children to a more costly school. In the case of charter schools, the supplementation is more likely to be from philanthropic contributions that increase spending in particular charter schools. In some cases, this financial augmentation from private sources to specific charter schools has been substantial (Baker, Libby, & Wiley, 2012) A larger voucher or greater charter school funding will allow more options in the marketplace with greater freedom of choice, possibly increasing productive efficiency as well as greater competition for students. But if families can add on to vouchers from their private resources as Milton Friedman (1962) proposed, there will be advantages for families with higher incomes and a potential decrease in equity. This is also a concern for charter schools if those with greater resources are able to attract more capable students.

Alternatively, the educational voucher can be differentiated by educational need by providing larger vouchers for those with handicaps and from poverty backgrounds to address equity. Schools will have greater incentives to attract educational needy students and provide the resources and programs to address their needs. Clearly, the financial arrangements can have profoundly different consequences for equity.

Regulation. Regulation encompasses the requirements set out by government for eligibility of schools to participate in the voucher or charter school system as well as guidelines for participation of families. Only schools that meet certain requirements will be eligible to redeem vouchers, and these requirements are designed to achieve certain policy goals. A particular source of contention is the eligibility for public funding of schools sponsored by religious entities. Although the U.S. Supreme Court approved the participation of religious schools in the case of the Cleveland voucher plan, many of the states have much stronger proscriptions against public funding for religious institutions than those in the U.S. Constitution (Zelman vs. Simmons-Harris, 2002).

Some voucher plans have emphasized a common curriculum and uniform testing as a condition of school participation. The purpose is to ensure that students are exposed to experiences that support social cohesion and that schools can be compared for their productive efficiency along common measures of student achievement. Some voucher analysts advocate that when schools have more applicants than places, applicants should be selected by lottery to provide fair access. This practice is typical for charter schools in many, though not all, states. Those who are more preoccupied with freedom to choose tend to favor government policies by which schools make all admissions decisions encompassing the premise that freedom of choice must also apply to schools in selecting the types of clientele that they prefer. Eligibility for vouchers may be restricted to certain populations in the name of equity. For example, public and private voucher programs in Milwaukee and Cleveland are limited to children from poorer families in order to give them choices outside of their neighborhoods. The Florida legislation limited vouchers to children in failing public schools. In each of these cases the regulatory aspects of design tend to favor some social goals over others. Many charter school plans also set preferences for serving such students.

Support Services. Support services can be used to enhance the effectiveness of the market in providing freedom of choice, productive efficiency, and equity. Competitive markets operate on the assumption that consumers will have access to a wide variety of choices as well as to useful information for selecting among them. In the United States, the availability of public transportation is very limited, necessitating a system of
school transportation from children’s neighborhoods to schools of choice. In the absence of school transportation, school choices and competition for students will be limited, reducing both the competitive efficiency of schools and creating inequities for those who cannot afford private transportation.

In order to make informed choices about the schools that they select for their children, families need good information. Accurate information on school programs and effectiveness as well as other important aspects of school philosophy and practice would need to be collected and disseminated to parents in appropriate forms to assist them in making decisions (Schneider et al., 2000). Schools have incentives to provide their own information through promotional materials and informational sessions for parents. However, there is little assurance that such information will be accurate and balanced, and it may be especially difficult for less-educated parents to process. Technical assistance might also be provided by government agencies through information and training to assist new schools and to advance the productivity of the entire sector. But comprehensive systems of information and effective systems of school transportation, though essential for extensive and informed choice, are very costly. In this respect, they will compete for resources that would otherwise be available to schools for instruction.

**Trade-offs in the Design of Different Voucher or Charter Plans**

The different provisions that are set out in an educational policy such as a voucher or charter plan reflect the diverse values and purposes for pursuing choice and privatization. More specifically, the application of the three policy tools of finance, regulation, and support services in designing a privatization or choice plan will determine its impact on choice, efficiency, equity, and social cohesion. In essence, each designer uses these policy tools to construct a plan that either implicitly or explicitly places greater weight on some goals rather than others.

**Designs for Freedom of Choice**

A plan that seeks to maximize choice would promote a wide range of school types; would provide either a large voucher or charter school funding for all students or a smaller basic voucher with provision for supplements from family resources or philanthropic contributions; would minimize the regulation of curriculum, admissions, and other dimensions of school operations; and would provide comparative information on schools and an adequate system of transportation. Such a design would give parents a large number of alternatives from which to choose. Much of this plan is attractive to persons who prefer to see the least government interference in the marketplace, especially if a voucher program offers a modest voucher and parents are able to add to it. But, libertarians may view the support services of information and transportation as unwarranted because of high cost and government intrusion, and they would favor private tuition supplementation by families rather than a large basic voucher from public funding or the establishment of government-supported charter schools.

**Designs for Efficiency**

Productive Efficiency is maximized when schools produce a given level and type of education for the least cost. This concept is somewhat difficult to assess, in part because under a system of freedom of choice, schools may be producing very different types of education, which makes it hard to compare outcomes across schools. The matching of educational offerings to the preferences of families in a competitive environment is viewed by market advocates as an efficient use of resources, even in the absence of comparative Efficiency measures that demonstrate that conclusion. Accordingly, designs that focus on Efficiency would ensure that the voucher or government support of charter schools is large enough (when combined with parental add-ons or philanthropic support) to attract many competitors into the marketplace. Regulations would be minimal because they would tend to inhibit competition. Some would argue that academic achievement is so central to the productivity of all schools that testing of students should be required and reported as a check on academic Efficiency.
Although support services such as information and transportation might increase competitive pressures on schools, the cost of those services would have to be taken into account relative to the effectiveness gains.

**Designs for Equity**

Equity in education refers to equality in access, resources, and educational outcomes for groups that have traditionally faced differences on these dimensions. From a finance perspective, an equitable design would seek compensatory vouchers where more funding was available for students with greater educational need such as those in educationally at-risk and handicapped categories or greater charter funding for such students. To ensure that income differences would be neutralized, families would not be permitted to add on to the voucher, although philanthropic funds might be permitted for the most educationally disadvantaged populations.

The most fundamental regulation determining equity relates to who is eligible to receive and use a voucher or a charter school. Thus far, the publicly funded voucher plans in the United States have been limited to students from lower-income families or those enrolled in failing schools. Equity-oriented regulations would also embrace nondiscrimination in admissions, with schools being required to choose some portion of their students by lottery in the event that there are more applicants than openings. Provisions encouraging or requiring that schools not limit themselves to a narrow social or ethnic population would also be needed in light of the evidence that peers have an important impact on educational outcomes (Zimmer & Toma, 2000). With respect to support services, transportation would be required to provide access to those who are less advantaged, and information would be needed for promoting informed parental choices among schools.

**Designs for Social Cohesion**

To promote social cohesion, the voucher would have to be large enough to provide a common educational experience for all students. It would have to be structured so that all students could gain access to schools where they would be exposed to peers from a variety of backgrounds. Such diversity in student composition means that parental add-ons to the voucher would probably be proscribed, because they would tend to place students from different income strata into different schools. Charter schools would also have to promote student diversity, as have magnet schools in large cities (Smrekar & Goldring, 1999). Regulations would focus on establishing common elements in curriculum and certain school activities including the possibility of all students engaging in community service. Support services might focus on the provision of technical assistance in helping schools develop a common educational core as well as the information and transportation to enable families to find and gain access to schools with heterogeneous student bodies.

**Incompatibilities and Trade-offs**

As these examples should make clear, no single system provides maximal results for all four objectives. Ultimately, the choice of design features will depend upon specific preferences and values as transmitted through democratic institutions. Those who place a high value on freedom of choice will probably be willing to sacrifice some equity and social cohesion provisions by eschewing regulations and support services and allowing parental add-ons to vouchers and promotion of philanthropic funding for charters. Conversely, those who place a high value on social cohesion will be willing to sacrifice some freedom of choice through establishing a common curriculum core and other standardized features of schools. Ultimately, much of the debate over the specifics of educational voucher plans revolves around the political power and preferences of the stakeholders.

This conclusion becomes obvious if one reviews the two most prominent early voucher plans. The seminal plan of Friedman (1962) focuses on freedom of choice and productive efficiency through heightened competition, arguably at the expense of equity and social cohesion. Friedman would provide a modest, flat voucher at public expense. Parents could add to the voucher out of private resources, and schools could set...
their own tuition. Regulation would be minimal, and there would be no provision for transportation and information. This approach would promote a very large number of alternatives at different levels of tuition, for those who could afford them, with few restrictions on schools that enter the marketplace, which would promote a large supply of alternatives. Clearly, social cohesion and equity goals would not be paramount.

Conversely, plans that emphasize social cohesion and equity tend to reduce freedom of choice through extensive regulation and may inhibit productive efficiency by establishing costly support services. For example, the Jencks plan (Center for the Study of Public Policy, 1970) would regulate admissions and curriculum and would require standardized testing and reporting of results (see also the proposal by Godwin & Kemerer, 2002). It would also provide larger (compensatory) vouchers for the poor and a system of transportation and information. Moreover, vouchers could not be augmented from private resources. Relative to the Friedman plan, a fixed-government voucher with no private augmentation would reduce freedom of choice. The high costs of providing information and transportation and monitoring the regulations for eligible schools would add considerably to the costs of the voucher system (Levin & Driver, 1997).

Although certain design provisions would improve outcomes for one or more goals, they typically would also reduce outcomes for other objectives. For example, provision of information and transportation would promote equity by improving choice options for those with the least access to information and transportation. But such provision would also raise considerably the noninstructional costs of the educational system, thereby reducing academic efficiency unless the academic gains from competition due to better information and access offset the high costs of transportation and information. Empirical studies of school competition suggest that this outcome is unlikely (Belfield & Levin, 2005b). Likewise, the establishment of regulations with continuous monitoring and enforcement could be used to increase equity and social cohesion, but at the sacrifice of freedom of choice and productive efficiency.

Advocates of vouchers may agree on the general case for vouchers or increased educational privatization, but may disagree profoundly on specifics. Strong differences arise even among persons with the same general political persuasion. Thus, many political liberals want to see greater freedom of choice for students in the inner city through educational vouchers, even though political liberals are usually viewed as antagonistic to market-based approaches for the delivery of government services. At the same time, cultural conservatives are deeply committed to a common curriculum and knowledge framework that should be required of all students and the schools where they are enrolled, which represents a substantial commitment to regulation (Hirsch, 1987). Political conservatives with libertarian views reject regulatory requirements entirely in favor of complete market accountability; that is, letting consumers decide what they want. The classic Friedman (1962) proposal is very close to this stance in its almost complete omission of regulation, despite his willingness to provide public funding for vouchers on the basis of the contribution of schools to a “stable and democratic society.”

Summary

The term educational privatization can thus mean different things depending on how it is applied and which dimensions of education are privatized. There are many different institutional forms of privatization, and even profoundly different versions of the same form, with potentially different educational consequences, and one can perceive parallel patterns among the different charter school arrangements among the states. The specific provisions for finance, regulation, and support services can create very different educational outcomes. Moreover, in many cases the improvement of one outcome through the use of these provisions will be accomplished at the expense of other outcomes. For all of these reasons it is best to address issues about educational privatization and school choice plans more generally by grounding them in the details of specific provisions such as those on finance, regulation, and support services. As a corollary, broad generalizations and conclusions about the impacts of educational privatization should be viewed with skepticism unless buttressed by supportive details and evidence.

Finally, it appears that claims of particular educational results for privatization and school choice such as charter schools are colored more heavily by ideological leanings than by supportive evidence. As the chapters in this section and other summaries of research on privatization demonstrate, the evidence on the consequences of the different forms of privatization and school choice is often mixed and contradictory (e.g.,
Belfield & Levin, 2005a; Gill et al., 2007; Zimmer, Guarino, & Buddin, 2010; CREDO, 2013; Zimmer & Bettinger, this volume; Bifulco & Bulkley, this volume). In this respect the ideological claims seem to have trumped the validity of the empirical evidence in the debate over educational privatization and choice (Belfield & Levin, 2005a; Hess, 2010).

Note

1. It is important to note that the present quest in school finance for addressing "adequacy of funding" is an equity concept. Adequacy does not mean that the level of funding must be the same for each child, since it generally assumes that the funding for each child will reflect differences in educational needs.

References


Charter Schools

ROBERT BIFULCO AND KATRINA BULKLEY

Introduction

Charter schools have been one of the most significant developments in U.S. education over the last 20 years. Since the first charter school program was introduced in Minnesota in 1991, 42 states and the District of Columbia have adopted charter school legislation. In the 2012–13 school year, roughly 6,000 charter schools were open nationally, enrolling more than 2.3 million students (National Alliance for Public Charter Schools, 2013). The growth of charter schools has continued steadily, with roughly twice as many students in attendance as listed in the 2008 edition of this volume. Although only about 4 percent of students nationwide attend charter schools, these schools serve more than 10 percent of the public school population in over 100, mostly urban, school districts (National Alliance for Public Charter Schools, 2012).

While state charter school laws vary considerably, charter schools are generally non-sectarian, free, public schools operating under a contract, or charter, granted by a public agency (the "authorizer" of the charter school). This contract usually lasts for a set number of years and must be renewed in order for the school to continue receiving public funding. In most cases, students are not enrolled in charter schools unless their parents apply for admission. Charters are usually open to all who wish to attend within a geographic area, are supported primarily by tax dollars, and are not allowed to charge tuition. Typically, oversubscribed charter schools are required to select students by lottery. Public funding for charter schools is tied more closely to enrollments than in the typical traditional public school, and charter schools operate with varying levels of independence from local school boards and district central offices.

The origins of the charter school concept can be traced to Ray Budde, an education professor, and to Albert Shanker, the former president of the American Federation of Teachers (Budde, 1988 and 1989; Shanker, 1988a and 1988b). As a result of complex politics and a changing landscape for educational reform, charter school proposals began to shift emphasis away from small-scale, teacher-led schools promoted by Budde and Shanker and towards external accountability, autonomy, innovation, and competition (for more about the politics of charter schools, see Bulkley, 2005; DeBray-Pelot, Lubinski, & Scott, 2007; Fuller, 2000; Henig, 2008). As a reform that is largely about school governance, no particular educational practices are associated with the charter school concept.

Multiple rationales have been offered for charter schools. At its core, the charter school approach involves schools that trade additional autonomy from state and local regulation for greater accountability to parents, through markets, and to government, through authorizers. The need to compete for students, advocates argue, pushes charter schools to be more innovative and higher quality than district-run public schools. By providing competition for traditional public schools, charter schools are also seen as a way to improve the achievement of all public school students (Kolderie, 1990; Nathan, 1996). In addition, advocates have argued that by breaking the link between where students live and where they go to school, choice programs allow poor and minority students trapped in racially and economically isolated neighborhoods to attend more-integrated schools (Viteritti, 1999). Others, however, have worried that charter school programs may increase segregation and the isolation of disadvantaged students (Cobb & Glass, 1999; Frankenber, Siegel-Hawley, & Want, 2010).

In this chapter, we review research that examines the effects of charter school policy on innovation and educational practice, student segregation across schools, and academic achievement. Policy questions have shifted over time from whether or not to establish charter schools to how to revise existing laws and policies to enhance charter school quality (Wohlstetter et al., 2013) and to how charter schools fit within broader systemic
Design and Implementation of Charter School Policies

Charter laws and practices have evolved over the past 20 years (Eckes, 2010; Shober, Manna, & Witte, 2006). Here, we review some of the key policy issues that emerged as states gained experience designing and implementing charter school programs.

Authorizer Roles and Practices

Many charter school laws identify establishing new forms of accountability as an explicit goal. Although competition for students is intended as one way to help make charter schools accountable to parents, the main factor that distinguishes charter schools from voucher programs is that they are also publicly accountable through the chartering process (Hassel, 1999). In principle, charter school contracts articulate clear goals and expectations with the understanding that renewal depends on a school meeting those expectations. The authorizers charged with approving, monitoring, and evaluating charters thus play a key role in ensuring public accountability for charter schools.

Authorizers vary by state and can include local educational agencies, state departments of education, universities, state charter school boards, and nonprofit organizations. Several authors have highlighted the challenges authorizers face in balancing accountability for processes and outcomes, support and oversight roles, and providing direction and allowing innovation (Bulkley, 2001; Fiske & Ladd, 2000). Research suggests that many authorizers lack the capacity to carry out their oversight functions in a meaningful way (Finnigan et al., 2004). A number of authors also note the difficulty of developing and enforcing clear expectations in educational contracts (Bulkley, 2001; Hannaway, 1999; Hassel & Batdorff, 2004).

The National Association of Charter School Authorizers has actively promoted standards for authorizing practices (Mead, 2012), but research into effective authorizer practices and policies is limited. In a study of Ohio charters, Zimmer, Gill, Attridge, and Obenauf (2013) find that students attending charters authorized by nonprofit organizations experienced, on average, lower achievement gains than students in charter schools authorized by government agencies. Although this study is unable to determine whether the achievement differences are due to the schools that nonprofits choose to authorize or the quality of support provided, and the results have yet to be replicated in other settings, these findings suggest that the effectiveness of different kinds of charter school authorizers merits greater attention.

Charter Schools and Systemic Reform

Most discussions view charter schools as providing competition for traditional public schools (for example, see Bettinger, 2005; Booker, Gilpatric, Gronberg, & Jansen, 2008; Rofes, 1998). Some observers, however, have suggested that charter schools and traditional public school districts should work more closely together (Bifulco & Reback, 2014). A "Portfolio Management Model" (PMM) imagines district central offices as overseeing schools governed in different ways, including privately managed schools and charter schools, with the main role for the central Office being to decide which schools warrant expansion or replication, and which should be closed (Bulkley et al., 2010; Hill & Campbell, 2012; Marsh, Strunk, & Bush, 2013). Advocates of PMM districts often envision charter or charter-like schools serving substantial proportions of students, which differs from the marginal role charters have played in most places to date.
Financing Issues

Public funding for charter schools is provided either directly from the state or from local school districts in the form of per-pupil payments. When payments are made directly by the state, charter school students are typically not counted in district enrollment for purposes of state aid, so that charter schools are at least partially financed by reductions in state aid to districts. Two key issues are how charter schools affect the fiscal condition of local school districts and the distribution of school funding across students.

Charter schools can affect a local school district’s fiscal condition by generating additional costs and/or by drawing funding away from the district. Arsen and Ni (2012) and Bifulco and Reback (2014) show that districts suffer negative fiscal impacts, at least in the short run, in places where enrollments are stagnant or shrinking, charter school students are a significant portion of total enrollment, and districts are required to make substantial per-pupil payments to charter schools. Bifulco and Reback (2014) suggest that, in these cases, states should design policies that can relieve the short-run burden charters place on districts without undermining long-term incentives to adjust to enrollment reductions, and that equitably distribute the burden of any additional costs that charter schools create.

Some proponents have raised concerns about the underfunding of charter schools. Charter schools often receive less than the average per-pupil spending in the local school district, and typically do not receive any extra funding for facilities (Speakman & Hassel, 2005). However, local school districts often have expenses and service responsibilities that charter schools do not have (Nelson, Muir, & Drown, 2003), and often provide in-kind services to charter schools such as facilities, transportation, health, and special education services (Nelson, Muir, & Drown, 2000; New York City Independent Budget Office, 2010). Differences in the characteristics of charter and traditional public school students further complicate comparisons of per-pupil spending amounts. For instance, charter schools typically serve fewer special education students and English language learners (CREDO, 2013). These issues not only complicate assessments of funding equity, but also raise questions for policymakers about how to set per-pupil payments and adjust those amounts for differences in student needs.

In addition to public funding, charter schools may receive funding from private sources. As in the case of traditional public schools, some charter schools may be better positioned to raise private resources, which raises questions about school funding equity (Reckhow, 2013). Some recent research suggests that charter school management organizations and private foundations provide considerable resources to some charter schools (see Baker & Ferris, 2011; Baker, Libby, & Wiley, 2012).

Looking Inside Charter Schools

Many charter school proponents emphasize charter schools’ potential for promoting innovation (Hassel, 1999; Nathan, 1996; Weitzel & Lubienski, 2010). Charter schools themselves represent an innovative approach to the way publicly funded schools are governed. Beyond this, however, advocates of charter schools hope they will foster innovation in the day-to-day operations of schools. Determining the extent to which charter schools foster innovation is complicated by the fact that “innovation” is difficult to define—for example, is innovation something completely new or new to a particular community (Lubienski, 2004)? A related motivation for charter school programs is to expand the range of programmatic offerings and educational environments from which parents and students can choose (Goldhaber, Guin, Henig, Hess, & Weiss, 2005). Such diversity, in theory, provides parents opportunities to match programs with their child’s needs and educators with opportunities to identify effective practices. This section examines research on practices within charter schools with an eye toward innovations and differences between charter and traditional public schools.

Charter Schools and New Organizational Structures

Charter school programs have spawned new kinds of organizational structures including education and charter management organizations (EMOs and CMOs). Originally, the charter idea focused on individual schools operating with autonomy from school districts. The challenges of “going it alone,” however, have motivated
many charter schools to partner with other groups and organizations. Such partnerships can provide charter schools with financial, human, political, and organizational resources (Wohlstetter & Smith, 2010). Although many charters partner with community-based and faith-based organizations, nonprofit CMOs and for-profit EMOs are the most prominent partners, with 36 percent of charters under an EMO or CMO umbrella in 2011–12 (Miron & Gulosino, 2013). CMOs and EMOs are organizations that include some form of “central office” that directly manages or supports multiple charter schools based on a shared mission and/or educational approach (Farrell, Wohlstetter, & Smith, 2012; Furgeson et al., 2012; Miron & Gulosino, 2013). While the EMO sector has experienced steady enrollment growth, the CMO sector almost doubled the number of students enrolled between the 2009–10 and 2011–12 school years (Miron & Gulosino, 2013); based on this growth and the attention CMOs have received in the political and foundation sectors, we focus the remainder of this section on those organizations.

The numbers vary based on the specific criteria used to identify an organization as a CMO, but one recent report counted 130 CMOs that managed four or more schools in the 2010–2011 school year, serving roughly 250,000 students (Furgeson et al., 2012). Farrell, Wohlstetter, and Smith (2012) identify multiple reasons for the growth of CMOs and CMO-managed schools, including: dissatisfaction with the outcomes and growth of stand-alone (non-CMO) charter schools; federal programs and funding to support replication of “high-quality” charter schools; authorizer practices; local political supports; and funding from nongovernmental sources. Growth has been spurred in part by substantial support from foundations. Quinn, Tompkins-Stange, and Meyerson (2013); Reckhow (2013); and Wohlstetter, Smith, Farrell, Hentschke, and Hirman (2011) consider support of CMOs to be a key example of “venture philanthropy” (Scott, 2009).

While CMOs offer a range of educational approaches, those linked with “No Excuses” models have received the most attention from policymakers, foundations, and the media. Such models, which are linked to the writings of Abigail and Stephen Thernstrom (2004) and Samuel Carter (2000), focus on measurable goals, increased instructional time, student comportment, selective hiring, and traditional approaches to teaching math and reading and often serve largely low-income students of color (Angrist et al., 2011; Dobbie, Fryer, & Fryer, 2011). KIPP (the Knowledge Is Power Program) and Achievement First are commonly identified “No Excuses” CMOs.

In their ongoing study of CMOs, Furgeson and his colleagues (2012) found that, compared to districts CMOs: are less likely to prescribe curriculum or behavior policies; provide more coaching and are more likely to use performance-based pay for teachers; and devote more time to instruction. The authors also identify several practices associated with positive impacts on achievement and attainment, including comprehensive behavior policies and intensive coaching of teachers.

Virtual or “cyber” and homeschool charters, collectively called nonclassroom-based charters, are another organization innovation that warrants attention (Huerta, d’Entremont, & Gonzalez, 2009). Cyber-charter schools rely primarily on self-administered computer-based instruction while maintaining responsibility for evaluating student performance. These schools have proven particularly attractive to homeschool parents because of the minimal role of government oversight and the central role of parents. According to Ahn (2011), cyber-charter schools are active in at least 17 states. In homeschool charters, parents more explicitly play a central role in designing activities and curriculum and assessing student performance. The growth of cyber-charter schools is occurring in parallel to a broader expansion of virtual education in public K–12 schooling, with K12 Inc as the dominant organization (Glass & Welner, 2011; Natale & Cook, 2012). Research has documented weak achievement and accountability in cyber-charter schools (Miron, Urschel, Aguilar, & Dailey, 2012; Molnar, 2013; Zimmer et al., 2009) and inappropriate use of finances in nonclassroom-based charters (Huerta et al., 2009; Miron & Urschel, 2012).

**Instructional Conditions**

The multifaceted conditions that shape instruction in schools and are sometimes tied to student achievement include curriculum and assessment, professional community, leadership practices, teacher expectations, and pedagogical approaches (Berends, Gold-ring, Stein, & Cravens, 2010). Evidence suggests that charter school educators experience greater autonomy around school decisions than those in traditional public schools, although the data on which much of that work is based predates the growing role of CMOs, which may reduce school level autonomy (Gawlik, 2007 and 2008; Renzulli, Parrott, & Beattie, 2011; Wohlstetter et al., 2013).
While some research has worked to look inside the “black box” of school practices, this is still an area in which further work is needed. Several studies examining the experiences of teachers and their working conditions, including the role of professional community, suggest that clear missions and instructional programs as well as time and support for collaboration may enable charter schools to foster positive working conditions for teachers (Bulkley & Hicks, 2005; Malloy & Wohlstetter, 2003). Despite the broad rhetoric of innovation, research suggests that practices in charter schools are often similar to those in traditional public schools (Cannata & Engel, 2012; Preston, Goldring, Berends, & Cannata, 2012; Weitzel & Lubienski, 2010).

**Teacher Staffing—Recruitment, Selection, Pay, and Retention**

Some have argued that freedom from bureaucratic constraints, and particularly union influence and collective bargaining agreements, may allow charter schools to adopt more effective teacher selection, assessment, and pay policies (Finn, Manno, & Vanourek, 2000; Hassel, 1999; Stuit & Smith, 2012). Evidence is mixed and inconclusive, but does not provide strong support for this argument.

Several studies find that charter schools tend to hire younger, less experienced teachers with fewer master’s degrees and more uncertified teachers than the typical public school (Burian-Fitzgerald, Luekens, & Strizek, 2004; Cannata & Penaloza, 2012; Carnoy, Jacobsen, Mishel, & Rothstein, 2006; Podgursky & Ballou, 2001). Studies also find that charter schools have higher rates of teacher turnover than traditional public schools (Burian-Fitzgerald, 2005; Podgursky & Ballou, 2001). Stuit and Smith (2012) find that lack of certification and experience, alongside low levels of union membership, help explain this high turnover.

Critics argue that schools with less experienced teachers and high-turnover rates are a problem because inexperienced teachers might provide inferior instruction, teacher turnover undermines instructional cohesiveness, and a high concentration of inexperienced teachers in charter schools reduces opportunities for mentoring by more experienced teachers (Carnoy et al., 2006). Kane, Rockoff, and Staiger (2008), however, suggest that high-turnover rates can be consistent with optimal teacher policies if schools are able to induce high-performing teachers to stay and low performers to leave. In an early study, Podgursky and Ballou (2001) found that charter schools are more likely to dismiss teachers for poor performance and to award merit-based bonuses—policies that may help to target retention to more effective teachers. A more recent study, however, raises doubts that charter schools have more effective retention policies. In a careful study of teacher attrition in Florida, Cowen and Winters (2013) find that teachers in charter schools are more likely to leave than teachers in traditional public schools, and that the least effective teachers, assessed using value-added measures of teacher performance, were no more likely to leave charter schools than traditional public schools.

Studies comparing other teacher qualifications have provided mixed results. Some studies have found that charter teachers are more likely to come from more selective colleges (Baker & Dickerson, 2006; Burian-Fitzgerald et al., 2004). An analysis that draws on more recent data, however, found that the qualifications of charter school teachers were similar to those of traditional public school teachers (Cannata & Penaloza, 2012). In a study of North Carolina teachers, Carruthers (2012) found that less qualified and less effective teachers, assessed using value-added measures, moved to charter schools—particularly when the teachers were leaving traditional public schools for charters that were in urban areas, were low-performing, or served higher proportions of nonwhite students.

**Effects on Segregation and Isolation**

The segregation of students by race, socioeconomic status, and ability is a long-standing policy concern. In this section, we discuss the effect of charter schools on two separate dimensions of segregation. The first is exposure that the typical student has to members of other groups, for instance, the exposure of the typical white student to nonwhite students, or the exposure of poor students to nonpoor students. The second is the isolation of students in schools with high concentrations of disadvantage, for instance, the percentage of poor students in schools that are predominantly poor.
Do Charter Schools Increase or Decrease Intergroup Exposure?

Some argue that by weakening the link between residential location and school attendance, charter schools have the potential to decrease school segregation, particularly in places with high levels of residential segregation and neighborhood-based school assignment policies (Gill, 2005; Greene, 2005). Others have worried that if different groups have different information or preferences about schools, then charter schools might increase segregation (Bifulco, Ladd, & Ross, 2009; Cobb & Glass, 1999; Frankenberg et al., 2010).

The challenge in estimating the impact of charter schools on segregation is to determine how much exposure students would have had to students from different backgrounds in the absence of charter schools. One approach has been to use longitudinal data to compare the compositions of the schools students attend before and after transferring to charter schools. This approach has important limitations. First, the prior school can be observed only for those students who transfer to a charter school from a traditional public school. Second, this approach ignores the fact that the student composition in a particular traditional public school might be influenced by the presence of charter schools, and thus might provide an inaccurate estimate of what charter school students would have experienced in the absence of charter schools. Nonetheless, studies that use this approach provide a better indication of the effect of charter schools on student exposure to other groups than do cross-sectional comparisons of aggregate or average measures of student composition across sectors.

Studies using longitudinal data have found that black students who transfer into charter schools typically transfer to a charter school with a substantially higher percentage of black students than the traditional public school they left behind. For instance, Bifulco and Ladd (2007) report that the average black charter school student in North Carolina transferred from a school that is 53 percent black to a charter school that is 72 percent black. A similar pattern has been found in studies of charter school students in other states and districts (Booker, Zimmer, & Buddin, 2005; Garcia, 2008; Weiher & Tedin, 2002; Zimmer et al., 2009). This finding, however, is not universal. In three places studied in this way—Chicago, Milwaukee, and Little Rock—black charter school students transferred into charter schools with lower percentages of black students than that of their previous traditional public school (Booker et al., 2005; Ritter, Jensen, Kisida, & Bowen, 2012). In all three of these districts, the average percent black in the prior traditional public school was very high—73 percent in Milwaukee, 78 percent in Little Rock, and 90 percent in Chicago. Together, these findings suggest that in settings where black students tend to have ample exposure to other racial groups, the introduction of a charter school tends to reduce that exposure. In settings where black students are already highly isolated, however, charter schools in some cases increase exposure to other racial groups.

Bifulco & Ladd (2007) also found that the average white charter school student in North Carolina transfers from a school that is 28 percent black to one that is 18 percent black. This pattern of white students transferring into charter schools with fewer minority students is seen in some other states, but not in others (Booker et al., 2005; Garcia, 2008; Weiher & Tedin, 2002; Zimmer et al., 2009). None of the studies that include a large enough sample of Hispanic charter school students have shown significant effects on the exposure of Hispanic students to whites, but in some cases, Hispanic charter school students experience increased exposure to black students. As Ritter, Jensen, Kisida, and Bowen (2012) point out, comparisons of this type tell us nothing about how charter schools affect the exposure rates of students who remain in traditional public schools. Few studies have examined this issue in any convincing way.

Most of the findings reported above are from places where the percentage of students attending charter schools at the time of the study was low. Thus, the effects of charter schools on overall measures of segregation and student exposure in these places are quite likely small. In places such as Washington, DC, and New Orleans, where charter schools serve a much higher proportion of students, the effects on segregation may be more pronounced. However, the choices made by parents who are drawn into charter schools as they become more widespread might differ from the choices made by parents who have selected into smaller charter school programs. Also, the existing evidence suggests that the effects of charter schools on exposure to other groups depend on current levels of exposure in traditional public schools. Thus, it is difficult to predict the effects of more widespread charter school programs on segregation.

Do Charter Schools Increase the Isolation of Disadvantaged Students?
Even if charter schools have little impact on the exposure of the typical student to other ethnic or socioeconomic groups, they could increase the percentage of students who attend schools with high concentrations of disadvantage. In this section, we refer to schools with high levels of racial isolation (e.g., more than 90 percent nonwhite), high shares of poor students or students from families with low levels of education, and/or very low percentages of high-achieving students, as "isolated" or "socially isolated" schools. Many have argued that concentrations of student problems and lack of resources in such schools is particularly detrimental for student development (Frankenberg, Lee, & Orfield, 2003; Kahlenberg, 2000).

Charter schools can increase the number of students attending isolated schools in two ways. First, evidence from different types of school choice programs shows that more-advantaged students, particularly students with college educated parents, are more likely to use school choice programs to opt out of their assigned public schools and are especially likely to opt out from assigned public schools that have concentrations of poor or low-achieving students. Such moves can leave students in those assigned schools especially isolated from more-advantaged, higher-achieving peers. Second, charter schools that exclusively or primarily target disadvantaged or low-achieving students may themselves provide intensely isolated environments. Although students might benefit from these types of charter schools if the school is able to tailor specialized interventions that address their particular needs, existing evidence that these types of charter schools benefit their students is thin. Thus, the worry that such schools limit the exposure of their students to more-advantaged and higher-achieving peers without offsetting benefits remains (Frankenberg et al., 2010; Garcia, 2008).

Direct evidence on how charter schools affect the racial and socioeconomic isolation of students who remain in traditional public schools is limited. Carnoy, Jacobsen, Mishel, and Rothstein (2005) and Henig and MacDonald (2002) show indirect evidence that some charters draw relatively advantaged students from schools with concentrations of minority and disadvantaged families. However, these studies provide no direct evidence on how charter schools affect the composition of traditional public schools or the numbers of students in intensely isolated schools. Using more direct evidence that identifies the schools that individual charter school students attended prior to transferring, Booker, Zimmer, and Buddin (2005) find that charter schools in Texas are generally not increasing the isolation of students in the schools from which they draw enrollments. Considering the strong preference it has given to authorizing charter schools for at-risk students, however, the Texas charter school program might be exceptional. Overall, then, it remains unclear whether charter schools typically draw the most advantaged students from schools with concentrations of disadvantage.

Booker, Zimmer, and Buddin’s (2005) findings for Texas, however, do suggest that by attracting particularly low-achieving students, charter schools themselves may be providing particularly isolated learning environments. A few studies have examined this issue more directly. Bifulco and Ladd (2007) find that over 30 percent of the charter schools in North Carolina served student populations that are more than 80 percent black, and that two-thirds of these schools had a higher percentage blacks than any traditional public school in the same district. Using data on a national sample of schools, Ritter et al. (2010) found that among students attending schools located in central cities, charter school students are 10 percent more likely than traditional public education students to attend intensely racially isolated schools.

In sum, it appears that the introduction of charter schools can increase the share of disadvantaged students who attend intensely isolated schools. However, it is less clear how much charter schools increase the share of students in isolated schools, under what policy conditions, and what the effect on those students who find themselves in more isolated environments might be.

Effects on Academic Achievement

Charter schools can influence student achievement by providing either higher-or lower-quality programs than those offered in traditional public schools, by helping to improve the match between educational programs and individual student needs and by influencing the achievement of students who remain in traditional public schools. In this section, we review the evidence on the effects of charter schools on the achievement of charter school students as measured by standardized tests. A small number of studies have also examined the effect of charter schools on important noncognitive outcomes, and we discuss findings from these studies as well.  

371
Key Questions

Much of the research on student achievement has focused on the average effect of charter schools on charter school students, as opposed to the effects of specific charter schools. Charter school programs are systemic interventions aimed at changing the institutional environment under which public schools operate, and the average effect on achievement of charter school students is important for evaluating these institutional changes. In addition to the average effect, it is also important to understand how effects vary by race, ethnicity, and socioeconomic status, across different settings, and by type of charter school.

If charter schools or some subset of them do have positive effects on student achievement, it is also important to understand why. If alternative governance and accountability mechanisms enable charter schools to improve student achievement, then that would suggest expanding the charter schools broadly. Alternatively, if the effectiveness of charter schools depends on features such as small school or class size, the appropriate policy response may be to replicate these features in traditional public schools. If charter schools are successful for reasons that cannot be easily replicated, such as favorable peer environments, then charter schools may not offer any lessons for system-wide improvement.

Methodological Challenges

Estimates of charter school impacts rely on comparisons of the achievement of charter school students to that of students who remain in traditional public schools. Because charter school students self-select, they are likely to differ in important ways, such as motivation, from students in nearby traditional public schools. The challenge is to identify a group of traditional public school students that provides a valid indication of what charter school students would have achieved had they remained in traditional public schools.

Early analyses of student achievement in charter schools, particularly those that focused on national samples, received extensive media coverage and were subject to contentious debates. Many of these early studies relied exclusively on statistical controls for observed student background characteristics to address potential selection bias. In the absence of measures to address unobserved differences between charter and traditional public school students, such methods are likely to provide biased estimates of charter school effects, and the direction of those biases is difficult to determine.

Two approaches to addressing potential unobserved differences between charter and traditional public school students have emerged. The first uses data that track individual students across years and allow one to control for individual students’ prior levels of achievement and test score growth. The second approach takes advantage of oversubscribed charter schools that use lotteries to determine admissions and rely on random assignment to eliminate selection bias. Both approaches have strengths and weaknesses, which are discussed further below.¹²

Student Fixed Effects Studies

When the achievement of individual students is observed at three or more different points in time, researchers can estimate models of achievement growth that control for individual student fixed effects. In such models, effect estimates are based on comparisons between the achievement gains made by students in charter schools and gains made by the same students in traditional public schools, effectively controlling for unobserved factors that have constant effects on achievement growth rates. Two studies that have tried to replicate estimates of the effect of choice schools derived from randomized assignment, have found that controlling for prior achievement using student level data greatly reduces the bias associated with methods that do not control for prior test scores (Bifulco, 2012; Fortson, Verbitsky-Savitz, Kopa, & Gleason, 2012). This method of estimating charter school effects also has the advantage that it can be applied to comprehensive samples of charter schools and thus produces estimates of average effects across large groups of charter schools.

Student fixed effect estimators have two drawbacks, however, both of which stem from the fact they rely on students who transfer between sectors. First, these estimators require the assumption that achievement growth rates vary over time in a similar way for students who transfer across sectors and those that remain in one
sector. If unusual difficulties in school prompt transfers across sectors, then this assumption may be violated (Hanushek, Kain, Rivkin, & Branch, 2007; Hoxby & Murarka, 2007). Second, even if the estimated effects of charter schools on students who switch sectors are accurate, the effects on this group of students may differ from the effects on students who are never observed in traditional public schools, and thus, might not represent the average charter school effect (Ballou, Teasley, & Zeidner, 2007).

Fixed effect analyses of charter school effects have been conducted in 14 states and large urban districts. The most consistent finding from studies of this kind is that newly established charter schools perform poorly (Bifulco & Ladd, 2006; Booker, Gilpatric, Gronberg, & Jansen, 2007; Hanushek et al., 2007; Ni & Rorrer, 2012; Sass, 2006). Even this finding, however, is not universal (Imberman, 2011; Zimmer et al., 2009). Evidence suggests that the average effects of more established charter schools on student achievement is less marked. While studies have found both negative and positive effects, average effect estimates from most studies are small and often limited to specific grade levels and/or subjects. Evidence suggests that these small average effects reflect wide variation across charter schools, some of which have large negative effects and others of which have substantial positive effects (Bifulco & Ladd, 2006; Hanushek et al., 2007).

Regarding variation of achievement effects across students, settings, and types of schools, it is difficult to discern consistent patterns. Most of the studies described above examine differences in charter schools by race/ethnicity, but consistent results do not emerge. Analysis by Hanushek, Kain, Rivkin, & Branch (2007) offers suggestive findings that charter schools in Texas have larger negative effects on students with lower initial achievement levels, and Imberman (2011) estimates that charter school have larger positive effects on girls than boys, but other studies have not examined variation in effects along these dimensions. Very few differences across types of charter schools have been examined, and the results are not consistent for those that have.

Two studies of charter school effects released by the Center for Research on Education Outcomes use student level, longitudinal data to estimate charter school effects in multiple states (CREDO, 2009). Despite the use of a different and less well understood method, the CREDO studies report results that are broadly consistent with the findings in other studies. In the more recent study, which included schools from 27 different states, 29 percent of charter schools showed significantly stronger achievement gains in math, 31 percent showed significantly weaker gains, and 40 percent showed math test score gains that were statistically indistinguishable from their matched traditional public school comparisons. Average effect estimates varied across states, ranging from a positive effect on annual student gains of 0.15 standard deviations in Rhode Island to negative effects of 0.19 standard deviations in Nevada.

A few recent studies examine student outcomes other than test scores. Imberman (2011) finds that although attendance at a charter school in an anonymous district had little effect on student test scores, it markedly increased attendance rates and decreased reported disciplinary incidents. Using data from Chicago and Florida, Booker et al. (2011) find that attendance at a charter high school markedly increases the probability of high school graduation, college attendance, college persistence, and earnings. These findings suggest that even where charter schools do not influence cognitive outcomes, they may improve noncognitive skills, which can be important determinants of later outcomes. These findings, however, are only suggestive. In the Imberman (2011) study, improvements in attendance and disciplinary incidents were not maintained by charter school students who returned to traditional public schools, suggesting that estimated effects might be due to policy differences between charter and traditional public schools rather than any change in noncognitive skills. Also, while evidence suggests that controlling for pretreatment test scores can substantially reduce bias in estimated effects on student achievement (Angrist et al., 2011; Bifulco, 2012; Fortson et al., 2012), it is less likely that pretreatment test scores can sufficiently reduce bias when estimating effects on high school completion decisions (Agodini & Dynarski, 2004).

**Lottery-Based Studies**

The second main approach to estimating the effects of attending a charter school uses data on students who participate in charter school admission lotteries. Effect estimates are based on comparisons of students who win lottery admissions and student who lose lottery admissions, and instrumental variables techniques are typically used to adjust effect estimates for the fact that some lottery winners do not attend charter schools,
while some lottery losers do. In principle, random assignment of admissions eliminates systematic differences between lottery winners and lottery losers, thereby eliminating potential selection biases.

Although random assignment of admissions is the most certain way to address selection bias, comparison of winners and losers of admission lotteries does not by itself ensure credible estimates of average charter school effects. The lottery process must produce truly random assignment of admissions, sample attrition must be low for both the lottery winners and losers, appropriate statistical models must be used to analyze the data, and results are more certain if most participants in the lottery comply with admission decisions. The most important limitation of lottery-based analysis is that they can be performed only for schools, grades, and years where applications exceed available seats by a sufficient amount, and evidence suggests that oversubscribed schools are a small and unrepresentative subset of all charter schools (Tuttle, Gleason, & Clark, 2012).

We analyzed eight studies that used data from admission lotteries to estimate charter school effects. Four of these studies examine the effects of three or fewer schools (Angrist, Dynarski, Kane, Pathak, & Walters, 2012; Dobbie & Fryer Jr, 2013; Dobbie et al., 2011; Hoxby & Rockoff, 2004; McClure, Strick, Jacob-Almeida, & Reicher, 2005). These say little about the effects of charter school programs per se and are better viewed either as demonstrations of how admission lottery might be used to estimate school effects or evaluations of particular schools.

Hoxby and Murkaka (2009) and Hoxby, Murkaka, and Kang (2009) report charter school effect estimates based on admission data from 43 of 47 New York City charter schools in operation at the time of the report. Estimates indicate that, on average, three to four years of attendance at one of these charter schools increases students’ third-grade achievement by 0.14 standard deviations in mathematics and 0.13 standard deviations in English. These estimates, however, are difficult to interpret. Even if all charter schools have admissions lotteries, schools that have lotteries in more grades, larger lotteries, or are oversubscribed to a greater extent are weighted more heavily. If schools with more applicants are more effective, estimates might overstate the average charter school impact. Also, attrition rates (i.e., the proportion of lottery applicants not observed in third grade) may be as high as 20 percent for some cohorts. If lottery losers who leave the study differ from lottery winners who leave, the estimated effects could be biased. These studies provide no analysis of attrition (Reardon, 2009).

Abdulkadiroğlu et al. (2011) estimate the effects of a set of five charter middle and four charter high schools in Boston. Estimates indicate that approximately one additional year of attendance in a charter middle school is associated with a 0.15 standard deviation increase in English language arts tests scores and a 0.40 standard deviation increase in math scores, and similar estimates for the high schools. These are substantial effects and larger than those obtained in most charter school studies. Angrist et al. (2013) find that these schools also had significant positive effects on indicators of college preparation and on college attendance.

Angrist, Pathak, and Walters (2011) extend these analyses to include a total of 22 of the 69 charter schools in Massachusetts serving middle and high school grades. They find that unlike the Boston charter schools included in the original study, charter schools in nonurban locations do not have positive effects on student achievement. In addition, the study provides nonexperimental estimates that control for prior achievement levels for the full set of Massachusetts charter middle and high schools serving traditional students. The results demonstrate that the nonexperimental estimates closely match the lottery-based estimates for the sample of charter schools for which lottery-based analyses are possible, and that the estimated effects for urban charter schools for which lottery analysis were not possible differ substantially from the estimated effects of schools in the lottery analysis. Particularly, nonlottery charter middle schools in urban areas showed small but statistically significant negative effects, and the nonlottery urban high schools showed positive, but significantly smaller effects than the lottery schools. These results suggest achievement effects in oversubscribed charter schools where lottery analyses are possible are not likely to be representative of all charters schools.

A study by Mathematica Policy Research reports lottery-based effect estimates for 36 schools across 15 states (Gleason, Clark, Tuttle, & Dwoyer, 2010). Estimates indicate that, on average, charter schools in this sample had small, negative effects that were for most measures of achievement statistically insignificant. The study also found that effects varied considerably across charter schools. In keeping with the results from Massachusetts, schools serving higher percentages of free-lunch eligible students showed statistically significant, positive effects on math scores after two years of attendance, as did schools serving students with
lower average baseline achievement, and urban schools. No subset of schools showed significantly positive impacts on reading achievement.

A separate study conducted by Mathematica Policy Research provides lottery-based effect estimates for 13 charter middle schools associated with KIPP (Tuttle et al., 2013). These estimates suggest that, on average, two years of attendance at a KIPP middle school increased scores on statewide math exams by 0.36 standard deviations. Effect estimates for reading were smaller and statistically insignificant. The results of this study are noteworthy for two reasons. First, in addition to the effects on state exams used for accountability purposes, the study also finds that these KIPP schools had positive effects on achievement measured by alternative tests that were not tied to state accountability programs and that included assessment of higher-order thinking skills. Second, like the Massachusetts study, this study demonstrates that nonexperimental analysis which controls for baseline test scores measured prior to charter school attendance closely matches estimates obtained from lottery-based analysis, and then uses these nonexperimental methods to obtain effect estimates for a broader sample of KIPP schools. These analyses, which include 41 of the 64 KIPP charter middle schools that had ever been in operation at the time of the study, suggest that KIPP schools, on average, have educationally meaningful positive effects across multiple subjects and in a majority of locations.

**Implications of Charter School Student Achievement Studies**

Although some positive achievement effects emerge in a few urban settings, most evidence suggests that charter schools have little or no effects on student achievement. The policy implications of these findings, and particularly the findings that certain charter schools tend to do well in urban settings, depend crucially on why these schools tend to increase student achievement levels.

The lottery-based studies indicate at least five ways that successful charter schools differ from both traditional public schools and other charter schools. First, the urban schools included in lottery-based studies in Massachusetts tended to serve student bodies with higher average baseline test scores than the traditional public schools that comparison group students attend (Abdulkadiroğlu et al., 2011; Angrist et al., 2011). This suggests that the composition of the student body may play a role in enabling these schools to create and maintain more favorable learning environments. Second, these charter schools tend to be smaller than other public schools. Third, both KIPP schools and the urban schools in the Massachusetts study tend to devote more resources to the educational process in the form of longer school days and longer school years than both traditional public and other charter schools, and at least one study has found that KIPP schools have substantially higher per-pupil revenues than comparison schools (Miron, Urschel, & Saxton, 2011). Fourth, given the observed characteristics of the teachers they employ and its connections with Teach for America, there is some reason to believe that KIPP schools are drawing teachers from different pools than are other schools. Fifth, all of the charter schools for which positive effects have been found report high levels of fidelity to the "No Excuses" approach to education.

Determining empirically which of these factors, or others, might account for achievement effects is challenging. Most of the studies that have tried, which are primarily the lottery-based studies, use small samples of schools, which makes it difficult to isolate the effects of a large set of potentially influential factors that are likely to be correlated and to interact in complicated ways. Thus, currently we do not know what factors account for the observed achievement effects. Nonetheless, policy implications depend on the answer to this question. If some urban charters are able to improve achievement because the selection of students enable them to create positive peer environments, any gains these schools achieve may be coming at the expense of losses at the schools these students would otherwise attend. If effects are due either to peer environments, additional resources, or drawing from a distinct and limited supply of teachers, then the success of a few charter schools may be difficult or at least very expensive to replicate more broadly. Or if the effects are due to extra resources or a "No Excuses" educational program, policy implications depend on whether these features would be as effective if adopted in other charter schools or by traditional public schools.

**Conclusion**
Wide variation across charter schools and states makes it difficult to draw broad conclusions about charter schools as a group or charter school policies. In addition, findings for charter schools operating as a small share of a much larger public school sector might not apply to a context in which charters provide a sizeable portion of the public school "seats" in specific districts. Nonetheless, the research to date allows us to draw some broad lessons for policymakers to consider, and points to critical areas for ongoing research.

Evidence suggests that innovative educational practices, a focus for early advocates, are not the norm across charter schools. Parental choice does not necessarily push towards academic rigor or innovative practices, perhaps because many parents seek more traditional practices (Stein, Goldring, & Cravens, 2011). Although not exclusive to charter schools, home and nonclassroom-based modes of instruction represent one potentially important innovation that has been prevalent among charter schools and deserves more attention from researchers. Several studies suggest that nonclassroom-based charters raise difficult accountability challenges. CMOs are an organizational innovation that have garnered substantial support from the political and foundation communities. Researchers need to look more closely at the practices and effects of CMO-managed schools and the ways in which different policy designs and authorizer engagement shape these schools.

Because charter schools still serve a small share of students in most places that have been studied, estimates of their effects on segregation and the social isolation of disadvantaged students have to date been small. Nonetheless, particularly in districts that have achieved high levels of integration and avoided creating schools with high concentrations of disadvantage, evidence suggests that broad reliance on charter schools could increase student segregation. Research into how policies can be used to avoid such outcomes, particularly in areas that rely broadly on charter schools and other forms of school choice, would be useful for policymakers.

Evidence suggests that, although some charter schools, particularly in urban areas, might help to improve student outcomes, charter schools as a group do not consistently do more than traditional public schools to increase academic achievement. A small number of recent studies suggest that some charter schools have had positive effects on other student outcomes, such as attendance, discipline, and educational attainment, suggesting important contributions to the development of noncognitive skills. Research on the effects of these types of outcomes is important because it can provide a more comprehensive assessment of the value of attending a charter school than studies of test scores alone. However, the data requirements and methodological challenges of such studies are considerable, and thus, evidence on other student outcomes is currently too thin to allow strong general conclusions about charter schools.

Given the existing evidence, and particularly the findings that charter schools do not consistently do more than other schools to increase student test scores, policymakers may want to revisit efforts to foster higher-performing charter schools through both supports and accountability. Additional research on charter policy and the practices of authorizers could enhance our understanding of strategies to support high-quality charter schools.

Finally, recent evidence suggests that some charter schools in urban areas may be achieving educationally meaningful improvements in student outcomes. However, the policy implications of these findings are far from clear and depend crucially on the reasons why these schools are able to achieve positive effects and whether the efficacious features of these schools can be replicated in traditional public or other charter schools. Identifying the key features of successful urban schools, testing the replicability and assessing the costs of those key features, and determining the importance of governance and accountability arrangements for promoting and maintaining effective schools are important issues for future research.

Notes

1. The "Race to the Top" and "Investing in Innovation" competitions put into place by the Obama administration have contributed to this growth (Wohlstetter, Smith, & Farrell, 2013).

2. For discussions of variations in many of the original state laws, see (Bierlein & Bateman, 1995; Scott & Barber, 2002; Wohlstetter, Wenning, & Briggs, 1995).

3. For recent accounts of the history and evolution of the charter school movement, see Lubienski and Weitzel (2010) and Wohlstetter, Smith, and Farrell (2013).
See Finnigan (2007) for an analysis of the autonomy available to charter schools.

Issues around competition between charter schools and traditional public schools are summarized by Gill and Booker in another chapter in this volume.

Although not discussed in this chapter, charters have also demonstrated other forms of organizational innovation such as alternative hierarchical structures, changing expectations for parental engagement, and new approaches to marketing (Angrist, Pathak, & Walters, 2011).

For a detailed discussion of both for-profit and nonprofit management organizations that manage charter and district-run schools, see (Miron & Gulosino, 2013). For a discussion of the history of management organizations in the charter sector and beyond, see Scott & DiMartino (2010).

Gains in instructional effectiveness from retaining only high-quality teachers can potentially outweigh losses in instructional effectiveness from relying more heavily on novice teachers.

Charter schools might also influence other aspects of segregation, such as the quality of intergroup relations and interactions within schools. (Gill, 2005). However, research on these other aspects is limited.

For a brief review see Bifulco, Ladd, and Ross (2009).

The effects of charter schools on the achievement of students who remain in traditional public schools are reviewed in the Gill and Booker chapter in this volume.

A few studies have used instrumental variable strategies as a complement to one of these approaches (Booker et al., 2011; Imberman, 2011). However, to date only a small number of studies have use this approach, and the validity of the instruments used is difficult to assess. The results of these studies are referenced below.

Several aspects of the analysis used are questionable; perhaps most noteworthy, is the decision to match on test scores observed after students had entered a charter school. In many contexts, matching on variables influenced by the treatment can result in biased impact estimates (Reardon, 2009; Fortson, Verbitsky-Savitz, Kopa, & Gleason, 2012).

Estimates of effects on achievement in grades four through eight are derived from regression models that include controls for prior year test scores that are themselves influenced by charter school attendance. Inclusion of control variables that are influenced by the treatment nullifies the benefits of random assignment. As a result, the analysis provides credible estimates only of the effect of charter school attendance on third-grade test scores. For a discussion see Reardon (2009).

References


377


Beyond the Rhetoric

Surveying the Evidence on Vouchers and Tax Credits

Ron Zimmer and Eric P. Bettinger

Introduction

Advocates for greater school choice have long hailed school vouchers as a mechanism for improving educational systems. These advocates argue that market incentives, created through the competition for students, can induce all schools to be more effective and efficient while creating greater educational opportunities for low-income families and, possibly, increasing racial and socioeconomic integration within schools.

Milton Friedman, who is largely credited with introducing the idea, originally proposed vouchers as a government-financed entitlement that students and their families could use to attend any school, including private schools (Friedman, 1955, 1962). Over time, support for voucher programs has made for strange bedfellows as conservative, pro-market advocates, Roman Catholic bishops hoping to reinvigorate parochial schools in urban cities, and more liberal-minded leaders of minority groups have advocated for the use of vouchers.

Voucher opponents, by contrast, argue that vouchers may exacerbate inequities in student outcomes. These opponents maintain that voucher programs can "cream skim" the best students and may increase inequities, especially if selection into voucher programs falls along racial or economic divides. Moreover, voucher programs reduce public school enrollments, thereby intensifying the fiscal strains already felt by public schools. Finally, these critics argue that vouchers can undermine the "common school" model, which instills common values for democratic citizens.

The debate over vouchers has heated up in recent years with the advent of publicly funded voucher programs having grown over time, including recently adopted statewide programs in Indiana, Louisiana, North Carolina, Ohio, and Wisconsin, in addition to dozens of privately funded programs across the country. Outside of the United States, voucher-like programs have increased in popularity; and while the motivation for these voucher programs differs from that in the United States, debates over these programs rage worldwide. In this chapter we examine many of the claims and counter-claims of the debate by surveying the research literature on both domestic and international programs.

While public funds in the form of grants and transportation subsidies have been provided to private schools in the United States for many years, we focus here on tuition support in the form of voucher payments, which are privately funded scholarships. Thus, this chapter represents a bridge between the research on educational choice and the research on private/public schooling differences. Researchers have long been interested in identifying the effects of private schools on student outcomes (e.g., Coleman, Hoffer, & Kilgore, 1982). However, in foreign countries, the distinctions between private and public schooling are often vague. In many cases, state governments provide substantial financial support and resources to private schools or to children who attend them (Toma, 1996).

We focus on differences in outcomes between students with and without vouchers. In some cases, researchers have extended beyond academic and behavioral outcomes or the distributional effects of voucher programs and examined the effects of voucher programs on residential housing markets (Nechyba, 2000), on subsequent public support for vouchers, or lack thereof (Brunner & Imazeki, 2006), and on winners and losers under voucher programs (Epple and Romano, 1998, 2003). We discuss these broader effects to the extent that
the studies shed light on the specific voucher programs that we review.

We have not included a review of research on tuition tax credits and deductions. That is not because they are insignificant—in fact tuition tax credits or deductions are actually more broadly employed than vouchers in the United States—but rather because they have been subject to little recent research. Researchers have yet to examine the achievement effects of education tax credits (ETCs). Instead, the existing research has demonstrated that ETCs typically lead to a decline in government tax revenue and that the subsidies are used disproportionately by families already sending their children to private schools. We refer readers to our previous literature review (2008) for a summary of tax credits/deductions and seminal work by Belfield and Wooten (2003).

Ours is not the first review of voucher programs. Gill et al., (2001), Gill et al., (2007), Levin (2002), McEwan (2000, 2004), and Zimmer, Guarino, and Buddin (2010) have all provided comprehensive literature reviews. Their reviews include the conceptualization of and rationale for voucher programs as well as the evidence of the efficacy of vouchers. This chapter builds on these reviews but focuses on the achievement and behavioral effects of vouchers on students and the effect these programs have on access and integration. For a review of the competitive effects of voucher programs, we refer readers to Gill and Booker (this volume). For a review of the cost effectiveness of voucher programs, we refer readers to Levin (2002).

Criteria for Inclusion of Studies

In recent years, a number of research studies have examined the effects of both publicly and privately funded voucher programs. For this literature review, we establish criteria for inclusion. Although these criteria are not exhaustive, they provide a framework for organizing and prioritizing lessons from the existing studies.

The first criterion for inclusion is the range of issues probed. The vast majority of the literature is limited to test scores, which makes it difficult to examine other outcomes. Although we highlight student achievement in this review, we also examine how vouchers affect high school graduation, college attendance, nonacademic outcomes, and labor market behavior. We also examine how they affect the distribution of students across schools by ability, race/ethnicity, and socioeconomic status.

Second, we pay particular attention to the assumptions required to elicit the causal effects of the voucher program. When examining an educational intervention, researchers try to identify the effects of the intervention by comparing the outcomes with the intervention to those that would have occurred without the intervention. In this case, once a student uses a voucher, we cannot observe the performance of the student without it. Though there are a number of methods for addressing this problem, each approach is based on one or more assumptions. We draw more attention to studies where the identifying assumptions seem most plausible and lead to the clearest picture of what would have happened in the absence of the program.

Third, we consider the quality of the data utilized for the analyses. Data on voucher programs are difficult to find. Some researchers expend significant resources attempting to track the applicants to a specific voucher program. Other studies rely on school-level data, drawing data from administrative records that may or may not fully represent the students in the voucher program. The ability of researchers to track individual students over time can greatly improve both the range of outcomes upon which it can focus and the likelihood that a study is informative.

Additionally, we note that attrition, while not necessarily a criterion for inclusion, can be a problem in these studies and needs to be considered when evaluating the quality of the research. High levels of attrition itself not only have important policy implications, but also can compromise the validity of the research. If the attrition is related to student outcomes, it can affect the ability of the researcher to identify the causal effects of the program.

Structure of Voucher Plans

Voucher programs differ greatly in their specific design. Much of this variation is related in part to the
differences in the motivation for vouchers among the voucher advocates (Levin, this volume). For example, the business community and many conservatives believe that vouchers can increase efficiency through increased competition. Thus, advocates with this perspective argue that voucher programs should have few regulations and restrictions, offer a flat modest voucher amount, and make no provision for transportation and information (e.g., Friedman 1955, 1962). By contrast, more liberal-minded voucher proponents argue that vouchers can promote equity by providing better educational opportunities to low-income families in inner-city schools. These proponents believe voucher programs should be means-tested, provide free transportation, require wide dissemination of information and strict regulation of admission policies, and require that state policymakers hold schools accountable for performance through testing of students (Center for the Study of Public Policy, 1970; Levin, 2002).

Ultimately, the differing motives of voucher advocates lead to large differences in the programmatic details. These details could affect who participates in a voucher program and thereby influence the generalizability of individual programs and studies. For instance, programs that restrict participation to low-income students are likely to have different distributional effects than does a universal voucher program. Also, programs that allow students to attend sectarian schools are likely to attract a particular set of students—specifically students seeking religious instruction opportunities. In addition, to the extent religious instruction affects the educational experience, this feature may affect outcomes such as student achievement, educational attainment, and other behaviors. Furthermore, in voucher programs that adversely affect the financial situation of public schools that lose students to private schools, such schools may be induced to improve their performance. Finally, the scale of the programs may affect whether these programs will generate general equilibrium effects that may be even larger than the localized effects of voucher programs (e.g., Nechyba 2000; Epple & Romano, 1998, 2003).

Because the programmatic design can affect outcomes, our review highlights the details of each voucher program and the associated outcomes individually. We synthesize the results in the final section and provide a summary of the programs and outcomes in Table 26.1 and Appendix Table 26.A.1.

Domestic Voucher Programs

Domestic voucher programs have been funded either by public taxes or private donations, primarily through philanthropists or foundations. While the scale and regulation of public and private programs may differ, they are similar in that they both provide a subsidy for students to attend private schools.

Publicly Funded Programs

The Milwaukee Voucher Program. The first large voucher program of note began in Milwaukee in 1990. With broad support from both liberals and conservatives, many advocates saw the Milwaukee voucher program as a way to extend greater educational opportunities to low-income students. As a result, vouchers were provided only to them. Initially, students could not attend sectarian schools, and available scholarships were capped at 1 percent of the total district population, but the cap has since been dropped and now over 24,000 students participate in the voucher program. From its inception, policymakers were interested in whether Milwaukee’s program improved academic achievement and educational opportunities for low-income minorities.

The first set of studies focused on the initial version of the voucher program. These studies were conducted by researchers at University of Wisconsin and Harvard and found conflicting results (Witte, 2000; Greene et al., 1997 and 1998). Much of these differences could be explained by differences in the control groups and led to a spirited debate. Hoping to add clarity to the debate, Princeton economist Cecilia Rouse (1998) used a strong research design that relied heavily on the fact that students who applied to oversubscribed schools were randomly selected for participation in the voucher program. Ultimately, Rouse found slight positive significant effects of the voucher program in math test scores, but none in reading.

More recently, a team of researchers led by Patrick Wolf matched individual voucher students to nonvoucher students to not only examine student achievement, but also educational attainment using high

386
school graduation rates. While matching strategies do not provide the same level of rigor as randomized design, some researchers have found that these matching strategies can replicate results from a randomized design (Bifulco, 2012).

Using the matching approach, the team released a series of reports. While the earliest years of reports did not find any significant student achievement effects for voucher programs, their last report, which followed students up to five years in the voucher program, found that when the voucher program was coupled with a new accountability program, voucher students experienced modest, but statistically significant positive effects in reading but no effects in math (Witte et al., 2012). In addition, the team found that voucher students, relative to nonvoucher students, had graduation rates four to seven percent higher; were slightly less likely to enroll in two-year technical postsecondary institutions, but were slightly more likely to enroll in four-year colleges (Cowen et al., 2012).

Policymakers also hoped that the Milwaukee experience would shed light on the impact voucher programs can have on access and racial integration. Research from the early years of the Milwaukee voucher program showed that the program generally attracted poor performing students from low-income families with a racial makeup similar to the district as a whole. A 1998–99 school year survey of voucher parents also showed that 62 and 13 percent of voucher students were African-American and Hispanic students, respectively, which was similar to the district population as a whole (Wisconsin Legislative Audit Bureau, 2000). However, these surveys also showed that the voucher parents had slightly higher educational levels (Rouse, 1998; Witte, 1996; Witte & Thorn, 1996) than did other parents in Milwaukee.

In a more recent study of the broader voucher program, a University of Arkansas team found that voucher participants were largely representative of the racial composition of the broader Milwaukee student community and that students switching because of the voucher program improved racial integration at the exiting school and reduced integration at the receiving school (Greene et al., 2010). In a separate study, Patrick Wolf and co-authors (2010) found that the voucher students were less likely to be special education students—special education students represent about 19 percent of all students district-wide relative to somewhere between 7.5 and 14.6 (depending upon the assumptions of their analysis) of voucher students. Overall, the voucher program, as required by the legislation, does serve low-income students, but among this population, the students in the voucher program are more advantaged.

The Cleveland Voucher Program. A second major voucher program was initiated in Cleveland in 1995. Similar to Milwaukee’s voucher program, support for Cleveland’s voucher came from both ends of the political spectrum. Cleveland’s program is means-tested and focuses on students from the Cleveland Municipal School District, one of the poorest districts in the nation. The program was initially intended to allow vouchers to be used at sectarian schools; however, court decisions immediately blocked this provision. Not until many years later were students allowed to choose among both sectarian and nonsectarian schools. Cleveland’s program now serves over 6,000 students.

Cleveland’s voucher program has not yielded clear answers about how the program affected student academic achievement. As in the Milwaukee program, the problem that arose with the evaluations was whether the researchers used an adequate comparison group. The primary evaluation (Metcalf et al., 2003) compared the academic achievement of voucher users to two groups of students: (1) students offered vouchers who did not use the vouchers, and (2) students who were not offered vouchers. For the second group, the researchers controlled for observable characteristics. Other researchers, including Gill et al. (2001), argued that the study failed to create an adequate control for students self-selecting into the program.

A later study by Clive Belfield (2006) tried to address this issue by comparing the performance of voucher students to students who applied but were rejected because they did not meet the eligibility requirements, including the level of family income. Belfield acknowledges the limitations of this comparison group, but he suggests that it is better than the alternatives. Overall, he finds little effect from the voucher program on academic achievement, but again, the inadequacies of the control for selection calls into question the reliability of the results.

Another key motivation for establishing the voucher programs was to improve students’ access to better schools. Kim Metcalf (1999) found that the vouchers were used primarily by minorities and students from low-income families. The mean income level of students utilizing the vouchers was $18,750, and these students were more likely to be African-American students than a random sample of Cleveland students. As in
Milwaukee, parents of voucher applicants had slightly higher levels of education (Metcalf, 1999; Peterson, Howell, & Greene, 1999). Therefore, like Milwaukee, the Cleveland program does serve low-income populations, as required by the legislation, but among this population, these students were more advantaged in terms of education.

**District of Columbia Program.** In 2004, The District of Columbia School Choice Incentive Act established the DC Opportunity Scholarship Program. Like the programs in Milwaukee and Cleveland, it is a means-tested program in which students with a family income below 185 percent of federal poverty line are eligible to participate. Students currently receive up to $8,000 each for grades K–8 and $12,000 for grades 9–12.

Commissioned by the U.S. Department of Education, a University of Arkansas team evaluated the program using random assignment from admission lotteries to create an unbiased treatment-control group comparison (Wolfe et al., 2010). Among the findings, the researchers found that voucher parents perceive their children to be attending higher-quality schools; 78 percent of voucher parents rated their child’s school as an A or B compared to 68 percent of parents of control students. The researchers found no statistically significant differences (at the 95 percent confidence level) in reading and math scores. The researchers also expanded the scope of the analysis by examining graduation rates. Here, they had to rely upon self reports with 63 percent of the target sample responding (Wolf et al., 2010, p. xx). The voucher parents reported a graduation rate of 82 percent compared to 70 percent reported by nonvoucher parents.

The lack of strong test score effects but strong graduation rates raises questions as to what is causing these effects, including whether the effects could be illusionary as schools could have lower graduation standards than traditional public schools. However, research on charter schools (Booker et al., 2011) has found not only similar graduation patterns, but higher levels of college persistence and earning levels for charter students (Booker et al., 2014) suggesting that these effects in charter schools could be real. If the same is true for private schools, the patterns suggest that the private schools may be offering advantages not fully captured by test scores.

Finally, in examining who is using the vouchers, the Arkansas team found that in the initial year of the program, the average family income of voucher users was $17,356 with only 6 percent of mothers of voucher students having a college degree (Wolf et al., 2007). Ninety percent of students were African-American, with an additional 9 percent Hispanic students. Seventeen percent were special need students. While these statistics obviously show that these students are from low-socioeconomic status families and are disproportionally minority students relative to the nation as a whole, the report provides no information on how these characteristics compare to those of the rest of the eligible population.

**Other Publicly Funded Voucher Programs.** Voucher programs are also in place in Vermont, Maine, Ohio, Wisconsin, Louisiana, Indiana, Colorado, and North Carolina. However, researchers have not yet examined how these programs affect student achievement.

**Domestic Privately Funded Programs**

Roman Catholic dioceses have provided scholarships to students to attend their schools for years. The use of philanthropists and foundation money for private school scholarships was first popularized by J. Patrick Rooney, Chairman of the Golden Rule Insurance Company of Indianapolis, who formed the Educational Choice Charitable Trust in 1991. This trust allocated private scholarships to low-income students to use at private schools in Indianapolis. Other programs quickly sprouted in Milwaukee, Atlanta, Denver, Detroit, Oklahoma City, and Washington, DC. The motivations for these programs vary by location. Some of the programs were established and financed by conservatives who wanted to either open up sectarian school options to more students or to create competition between private and public schools. Other programs (e.g., Oklahoma City) were funded by liberal voucher proponents who wanted to provide additional schooling opportunities for inner-city students.

In 1994, the Walton Family Foundation threw its support behind the movement by providing a $2 million grant to form Children’s Educational Opportunity Foundation (CEO America) to support and create existing and new scholarship programs. In the next few years, the number of privately funded programs quickly expanded to dozens of urban areas across the country. Of these programs, one of the most ambitious was a CEO America program in San Antonio funded at $5 million a year for at least 10 years. The program provided...
full scholarships to over 14,000 “at-risk” students, making it the largest privately funded voucher program in existence at the time.

In 1998, after the rapid proliferation of these programs, the Walton Family Foundation, through a $100 million grant, formed the Children’s Scholarship Fund (CSF), which went beyond giving scholarships to any one particular location. Rather, the program partnered with local funders to provide scholarships in cities across the nation. Due to high demand for these scholarships, in 1999 Walton provided an additional $30 million in funding, and in April of that year 1.25 million low-income students applied for 40,000 partial scholarships. Over the last decade, more and more of the programs developed, often receiving very little national attention, making it difficult to know exactly how many of the programs currently exist.

Of these programs, CSF has received the most national attention partially because of its relatively large scale and partially because its method of assigning private school scholarships is conducive to rigorous research utilizing randomization methods. To distribute scholarships equitably, CSF uses lotteries. The use of lotteries allowed researchers to use the randomized designs, that gold standard of evaluative research, and to compare students who participated in the lotteries to determine the effectiveness of these programs. Such studies have been carried out in Charlotte, Dayton, New York, and Washington, DC, among other locations. For each program, scholarship recipients and applicants in the city programs completed questionnaires and took the Iowa Test of Basic Skills (ITBS) at baseline and at annual follow-up sessions. While Greene (2000) did find a positive overall effect in Charlotte, these studies have generally shown no effect for the overall student population. They have, however, shown some positive effects for African-American students, who made up a small portion of the overall voucher recipients (Howell & Peterson, 2002; Myers et al., 2002). However, these results are not without controversy.

In 2002, Princeton economists Alan Krueger and Pei Zhu reanalyzed the data from the study of New York’s CSF program and found an insignificant effect for African-Americans. Arguing that Howell and Peterson (2002) and Myers et al. (2002) excluded too many students from their analysis, Krueger and Zhu expanded the analysis by including students with missing baseline test scores. Krueger and Zhu (2004) also redefined who was African-American. While the original analysis identified a student’s race by the race of her mother, Krueger and Zhu identify as African-American students whose mother or father are identified as African-American. This more inclusive definition results in a scholarship effect size that is smaller than reported in the original study and is not statistically significant. Still later, in a paper designed to develop ways to deal with noncompliance of treatment, John Barnard and colleagues (2003) found positive and significant math effects for students who applied from low-performing schools, and these results were strongest for African-American students—results largely consistent with the original set of results from the original set of studies.

More recently, Matthew Chingos and Paul Peterson (2012) revisited the New York scholarship program, but this time examined educational attainment in the form of college attendance. As in the original set of studies, the authors take advantage of the random assignment of students and found that African-American students were 20 percent more likely to attend college, 25 percent more likely to attend full time, and 130 percent more likely to attend a selective college. These results are impressive, but again, these results hold only for a small subset of the original study population.

Taking a different perspective, economists Bettinger and Slonim (2006) examined the impact of the CSF program in Toledo, Ohio, on nonacademic outcomes. Before beginning their data collection, the researchers interviewed principals at schools where students could have used the voucher. They compiled a list of outcomes that educators indicated they were trying to influence. Using surveys and new methods from experimental economics, they attempted to measure the effects of the vouchers on traditional outcomes such as test scores and on nontraditional outcomes such as student and parent levels of altruism, patience, and self-confidence. To test the effects of the voucher program on altruism, Bettinger and Slonim gave each student $10 and invited them to share some of their money with charities. In this context, they show that voucher recipients gave more to charities as a result of the voucher program.

While the achievement effects of voucher programs remain controversial, the types of students these programs serve are quite clear. Given that these programs were designed to target impoverished students, it is not surprising to find that participants have relatively low income. Peterson (1998), for example, found that the average income level of families of students participating in the New York scholarship program was only $10,000. The parents’ education level of the participants, however, was slightly higher than the average level of
the eligible population. In Washington and Dayton, Howell and Peterson (2000) and Wolf et al. (2000) found that the families of participating students had an average income level of $18,000. In terms of race, 100 percent of students participating in the Washington, DC, scholarship program were minorities, while 95 percent of the students in New York were minority, mostly Hispanic students. In Dayton, the percentage of minorities was lower but was still 66 percent. This research also suggests that the participating students were well below the national averages in test scores (Wolf et al., 2000).

Finally, research has also found that these programs served a low percentage of special education students. For example, in New York, 9 percent of participating students had disabilities compared to the district-wide average of 14 percent of special education students (Myers et al., 2000). In Charlotte, 4 percent of participating students had disabilities, while the district-wide average of special education students was 11 percent (Greene, 2000). The situation differs somewhat in Washington, where 11 percent of students were disabled compared to district-wide averages of 11 percent special education students (Wolf et al., 2000). In addition, the students using the vouchers may have less severe disabilities that are not taken into account in these comparisons.

While the research of who is taking advantage of these private scholarship programs raises some concerns of whether these voucher programs are fully inclusive of special education students, it shows that the programs have in general served disadvantaged students as intended.

International Voucher Programs

There are a number of educational voucher or voucher-like programs across the world, including programs in Chile, Colombia, Sweden, Netherlands, Belize, Japan, Canada, India, and Poland. These voucher programs often differ significantly from the U.S. programs either in terms of the motivating factors or the policy context. For example, many of the non-U.S. programs are motivated by the goal of increasing school attendance among girls or low-income students. Also, the relationship between church-sponsored private schools and public schools is less defined in other countries where, in many cases, religious groups operate public schools.

We discuss three voucher programs. Two of the programs, Colombia’s PACES voucher program and Chile’s national voucher program have been studied extensively and were the result of ambitious national policies. We also discuss a recent voucher program designed by researchers in India.

Colombia PACES Program

One of the largest voucher programs in the world operated in Colombia from 1992 to 1997. The Programa de Ampliación de Cobertura de la Educación Secundaria (PACES) offered educational vouchers to over 144,000 low-income students throughout Colombia. At the onset, the voucher program was fairly generous in that it awarded full private school tuition to secondary school students who wished to transfer from public to private school at the start of their secondary school experience. Over time, the value of the voucher did not keep up with the cost of inflation.

In contrast to the United States, where policymakers or private foundations espouse vouchers as a means of generating more competition between schools or as a way of providing alternatives to disadvantaged students in urban schools, policymakers in Colombia used vouchers to expand the capacity of the public school system. Most high school buildings host multiple high school sessions per day. For instance, one high school might meet in the early morning, another around midday, and yet another into the evening. Because most private schools were not overcrowded, Colombia established the voucher program as a means of taking advantage of excess capacity in the private system.

One of the unique features of Colombia’s PACES program was its use of randomization in the assignment of vouchers. When the program was oversubscribed, each municipality held a lottery to determine which applicants received the vouchers. Researchers have made use of these lotteries to identify the effects of the educational vouchers by comparing voucher lottery winners and unsuccessful lottery applicants. The strength of this approach is that it provides convincing causal evidence of the effects of the voucher on the type of students who apply for a voucher. Whether these effects would generalize to other populations of students is
Using this approach, Angrist, Bettinger, Bloom, Kremer, and King (2002) find that within three years voucher students had completed about 0.1 school years more than their peers and had test scores about 0.2 standard deviations higher than those who did not receive vouchers. They also find that the incidence of child labor and teenage marriage was lower as a result of the educational voucher. Based on follow-up research on students’ high school careers, Angrist, Bettinger, and Kremer (2006) find that voucher lottery winners were 20 percent more likely to graduate from high school and had more behavioral problems. Although the voucher winners had access to fewer resources in the schools they attended than did the voucher losers, voucher winners in this group had better academic outcomes.

Work by Bettinger, Kremer, Kugler, Medina, Posso, and Saavedra (2014) extends the evaluation of PACES to tertiary education outcomes and to labor market outcomes. The authors use administrative records to track students throughout their college careers and their initial years in the labor market. They find that educational attainment among voucher lottery winners is 0.2 years higher than among lottery losers. They also find that wages are 5–10 percentage points higher among voucher lottery winners and that voucher lottery winners are less likely to have been married as a teenager or to have a spouse who was a teenage parent.

One explanation for the better outcomes is that they reflect changes in student behavior associated with the incentives provided to the students. Students kept their voucher only if they were successfully promoted to the next grade (Angrist et al., 2002). Repeating a grade is fairly common in Colombia, and based on the behavior of voucher lottery losers, voucher winners who repeated a grade would likely have had to enter the workforce prematurely. The prospect of keeping the voucher and avoiding the labor market may have been enough incentive to encourage voucher winners to work harder than other students. Bettinger, Kremer, and Saavedra (2010) argue that this is a plausible mechanism especially since voucher effects are present even among a subsample of voucher winners who, as a result of some odd features of PACES voucher program, attended schools with less academic quality than those attended by voucher lottery losers. Regardless of the mechanism, the vouchers put students on trajectories in their teenage years which allowed them to attain higher levels of schooling and higher-paying jobs as adults.

In Colombia, the evidence on the impacts of stratification is less developed and conclusive than the effects on achievement. Colombia’s educational vouchers targeted low-income families living in the poorest neighborhoods, and Ribero and Tenjo (1997) report that the vouchers were largely effective in reaching this population. However, applicants to Colombia’s voucher program did not constitute a random sample of families from these neighborhoods. Voucher applicants came from families with higher educational levels than other families in the same neighborhoods (Angrist et al., 2002).

The Colombian voucher program was instituted to increase students’ access to schools. No systematic research has evaluated whether Colombia’s voucher program has actually improved access to education; as a result, it is unclear whether the voucher program increased overall enrollments among the most disadvantaged families.

**Chile**

In 1980, Chile embarked on an ambitious series of educational reforms designed to decentralize and privatize education. At the urging of Milton Friedman, who along with other prominent economists advised the Pinochet government at the time (Rounds, 1996), Chile established perhaps the world’s largest ongoing voucher program. The program offered tuition subsidies to private schools. Only a small number of elite private schools did not accept the voucher, and new schools entered the educational marketplace in response to the voucher program.

The true value of the voucher has fluctuated not only with business cycles and inflation, but also by school. Prior to the reform, the federal government had subsidized many private schools by giving them about half of the per-pupil amount that was spent on government-controlled schools. After the reform, these schools received the same amount as the public schools (Hsieh & Urquiola, 2003). Moreover, as part of the reform, the government decentralized public schooling. Instead of managing all expenses for public schools, the central government ceded control and began giving public schools a per student subsidy. Prior to the reform, the budgets of public schools were insensitive to enrollment; however, after the reform, local public schools lost
money when students transferred to voucher schools.

Unlike Colombia, where vouchers targeted poor students, the Chilean system was available to all students. Additionally, unlike other programs that do not allow selective admission, voucher schools in Chile could admit the students they most preferred. As a result of these policies, Rounds (1996) found that the poorest families were less likely than other families to enroll their children in voucher schools.

Because of the nature of this admission policy, research on the efficacy of Chile’s voucher program has been much more difficult to interpret than the Colombian research. Some of the early evidence suggested that voucher schools modestly outperformed public schools. This finding was common in many papers (e.g., Bravo, Contreras, & Sanhueza, 1999; McEwan & Carnoy, 2000) but was sensitive to the types of controls included in the empirical model, the specific municipalities included in the sample, and the statistical methods used. McEwan (2001), for example, found that Catholic voucher schools tended to be more effective and productive than other schools, but only for certain specifications of the model. McEwan and Carnoy (2000) show that academic achievement is slightly lower in nonreligious voucher schools, particularly when located outside the capital. Given that these voucher schools have less funding, however, McEwan and Carnoy suggest that they could still be more cost-effective than public schools.

As research on the Chilean system continued, many researchers took note of the fact that the voucher program altered the composition of both public and private schools. For example, Hsieh and Urquiola (2003) show that after the reform began in 1981 the drop in public school enrollments was accompanied by a reduction in the average family income and the average achievement of the students who remained in the public schools. Thus, it was the more affluent families who transferred their children to the private sector following the reform. Hsieh and Urquiola suggest that this shift in student populations could account for the finding that private schools appear to be more effective than public schools. McEwan and Carnoy (2000) and Carnoy (1998) offer similar evidence that parents in voucher schools have higher incomes and greater levels of education than parents in public schools. In the early 1990s, many voucher schools began charging tuition in addition to the voucher, and the difference in parents’ incomes and education levels between these tuition-charging voucher schools and the other voucher schools was significant (Anand, Mizala, & Repetto, 2006).

How this increased sorting across voucher and nonvoucher schools affects student achievement depends on the nature of peer effects. If improvements in peer quality lead to better educational outcomes for voucher users, then the sorting associated with the voucher program could improve their outcomes. At the same time, given the exit of high-quality students from the public schools, the students remaining in those schools may have systematically worse outcomes because of the greater disadvantage of their peers. The aggregate effect of the voucher depends on the strength of these two effects.

Hsieh and Urquiola (2003) argue that the only way to identify the overall effects of the voucher program is to focus on aggregate outcomes because it is difficult, if not impossible, to remove the selection bias inherent in comparisons of different schools. The change in aggregate test scores reflects both the direct achievement effects for voucher recipients and the indirect effects for students who remain in public schools. When the authors look at changes in aggregate test scores throughout Chile, they find no evidence that the voucher program increased overall test scores. (See Gill & Booker, this volume, for additional discussion of this research.)

Subsequent research by Gallego (2005) provides new evidence on the Chilean voucher program. Gallego notes that the number of Catholic priests existing in a community in the early 1950s can predict the number of and overall enrollment in voucher schools. This relationship arises because most voucher schools were affiliated with Catholicism both before and after the reform. Gallego (2005) uses this relationship to identify the effects of the voucher program. Specifically, he uses the number of priests as an instrument for the penetration of the voucher program in a specific market. His findings suggest positive effects of the voucher program on the academic outcomes of students throughout municipalities where the voucher program had more penetration. While the result may be indicative of competitive effects, it is driven in part by the effects of the program on voucher recipients. It echoes earlier research (McEwan, 2001; Carnoy & McEwan, 2000) which suggested that voucher schools affiliated with Catholicism had better outcomes than other voucher schools, public or private. As Carnoy and McEwan (2000) show, Catholic schools produced better students at a lower cost than other public or private voucher schools.

In 2008, the Chilean government modified the voucher program by increasing the funding by 50 to 60
percent for students from low-socioeconomic families. The increased subsidy was conditional on schools altering their admission policies and accepting greater accountability, and 68.4 percent of schools signed up for the new voucher program (Carrasco, 2013). Low-income students who participated in the new voucher program experienced a 0.18 standard deviation increase in math test scores and a 0.07 standard deviation increase in reading scores (Carrasco, 2013), but it is not clear whether these effects are the result of the voucher or more spending.

India

From 2008 to 2012, the Andhra Pradesh School Choice Experiment operated in five districts throughout the state of Andhra Pradesh. The voucher experiment used randomization both at the village level and within villages. Villages signed up to be part of the voucher program and conducted sign-ups for the voucher lottery within their village. After these prospective sign-ups were complete, villages were selected at random to participate in the voucher program. If admitted to the voucher program, the village then allocated vouchers using a lottery among those who had previously signed up to participate.

The resulting experiment created multiple groups of interest. Within villages, voucher lottery winners and losers could be compared as in prior studies. Researchers could also compare nonvoucher applicants across villages to detect whether the voucher affected the overall school system leading to improved performance for all students. Muralidharan and Sundararaman (2013) show that within villages voucher lottery winners had test scores that were 0.13 standard deviations higher than those of voucher lottery losers. However, they find no test score differences across villages, suggesting that the voucher program did not generate systemic changes in the school system. Given that schools knew that this was a temporary program, it is not clear that there were incentives to make large changes.

Summary and Conclusion

Table 26.1 summarizes the achievement findings from both domestic and international voucher programs. It highlights the location, program, methods used in the analysis, and their key findings.

Advocates for school choice often cast voucher and voucher-like programs as a possible silver bullet to improve educational performance, both within the United States and

Table 26.1 Summary of Domestic and International Voucher Programs

<table>
<thead>
<tr>
<th>Voucher site</th>
<th>Authors</th>
<th>Measures/controls</th>
<th>Identification strategy</th>
<th>Key results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milwaukee, WI</td>
<td>Witte (2008)</td>
<td>Voucher users</td>
<td>Regression with controls</td>
<td>No substantial effect; significant effects in both reading and math</td>
</tr>
<tr>
<td></td>
<td>Greene, Petrou, &amp; Do (1997, 1998)</td>
<td>Voucher users</td>
<td>Quasi-experimental design with covariate controls</td>
<td>Significant effects in math only and the effects grow over time</td>
</tr>
<tr>
<td>Rosier (1998)</td>
<td>Voucher users</td>
<td>Quasi-experimental design with covariate controls</td>
<td>Significant effect in reading only and only after five years of following students; the effect may be due to the size of the voucher program coupled with an accountability program started in the fifth year; positive effects for graduation rates</td>
<td></td>
</tr>
<tr>
<td>Witte et al. (2013)</td>
<td>Voucher users</td>
<td>Quasi-experimental design with covariate controls</td>
<td>No effect</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muralidharan &amp; Sundararaman (2013)</td>
<td>Voucher users</td>
<td>Observation with covariate controls</td>
<td>Positive effects for graduation rates</td>
</tr>
<tr>
<td>Cleveland, OH</td>
<td>Cleveland et al. (2003)</td>
<td>Voucher users</td>
<td>Quasi-experimental design with covariate controls</td>
<td>Little effect from the voucher program</td>
</tr>
<tr>
<td>Buffalo (2006)</td>
<td>Voucher users</td>
<td>Observation with covariate controls</td>
<td>Negative effects</td>
<td></td>
</tr>
</tbody>
</table>

393
abroad. While there certainly are cases in which voucher programs have led to impressive results, the results have not always been positive and have not consistently lived up to the expectations promised by voucher advocates. As Rick Hess points out in an interesting *National Affairs* piece (2010), these advocates may have set expectations too high—providing ammunition for opponents and deflating the morale of the advocates. Also, as pointed out by Hess and others (Levin & Belfield, 2003), part of the explanation for the inconsistent results across the voucher programs could be the variation in the voucher policies themselves, which could affect the ultimate success of these programs.

With respect to who is taking advantage of voucher programs, the research generally shows that students are low-income students, especially in voucher programs that are means-tested and that specifically target such students. However, among this population eligible for the vouchers, the voucher students generally have more-educated mothers and are less likely than nonparticipants to be special education students. In general, researchers have done a relatively poor job of examining how voucher programs affect the racial and economic integration of schools. Of interest to policymakers is not only the mix of students in the private schools chosen by voucher recipients, but whether the schools are more or less integrated than the schools that the voucher students previously attended.

To shed light on this issue, researchers need to use longitudinal student-level data so that they can track the movement of individual students and explicitly compare the mix of students in the schools they leave and enter similar to what has been done in the charter research (Booker et al., 2005; Bifulco & Ladd, 2007; Zimmer et al., 2009; Zimmer et al., 2011). In doing so, researchers could also examine more explicitly whether voucher programs are "cream skimming" the more motivated or advantaged students within the targeted group. In addition, because voucher programs are typically small in scale, it is hard to generalize to large scale programs. For a program with massive movements of students, an important policy consideration is not only who is participating in these programs, but what happens to the racial/ethnic and ability distribution of students in public schools.

In sum, while new research is emerging, the research community has yet to come to consistent conclusions about the merit of voucher and voucher-like programs. Researchers have also failed to provide strong insights into the reasons why voucher programs may be more or less successful across locations. Previously, Levin...
Levin and Belfield (2003), and Gill and colleagues (2007) argued that policy context matters. To get at this issue, researchers may need to design grander experiments as suggested by a National Research report (Ladd & Hansen, 1999). Such experiments may need to be large scale or cut across multiple sites in which the design would allow variation in the voucher amounts, access to information, transportation provisions, and participation constraints. By varying these policy levers, we can gain a greater understanding of how policies affect outcomes. But even such studies are unlikely to shed light on a number of key policy issues related to vouchers, such as how they affect the fabric of the overall education system.

Also, future research needs to peer inside the "black box" and to examine the mechanism through which outcomes differ among schools. Voucher researchers should take the opportunity to look at private schools that may be doing things differently and how the variation in the operation affects outcomes. In addition, because results for non-test score outcomes such as behavioral and educational attainment appear to be more promising, we encourage researchers to continue to explore these outcomes to see if these results are broadly true across a number of locations.

In addition, it would be useful to have more research on the costs of transitioning to a voucher program including costs of gathering new information, record keeping, and providing transportation. One study (Levin & Driver, 1997) shows that these costs may not be trivial.

Finally, it is also interesting to consider the value of voucher programs relative to other forms of choice. When the idea of vouchers was first introduced, the only choices a family could make were sending their students to private schools without subsidies or choosing among various public schools based on residential location. However, over the last 60 years, a number of other alternatives have evolved including charter schools, magnet schools, and other interdistrict and intradistrict choice programs. A relevant policy issue is the advantages and disadvantages of voucher programs relative to these other choices, especially when many of these options are politically more feasible. In sum, we have much to learn about vouchers as a policy tool.

Notes

1. While school vouchers were originally envisioned to be certificates of entitlements, the programs that fall under the domain of vouchers have expanded over time. These include privately funded scholarships and educational tax credits. In general, throughout the document, we use the word voucher when we are referring to these various programs.

2. A number of voucher studies focus on parental satisfaction among voucher schools (Howell & Peterson, 2002; Greene & Hall, 2001; Peterson, Campbell, & West, 2001). This literature is difficult to interpret since parental satisfaction is a significant predictor as to whether a family pursues an educational voucher. Not surprisingly, the general finding is that parents whose children are using the voucher tend to have higher levels of satisfaction than parents in the public sector.

3. Indeed, a large literature in economics has explored whether religious schools improve student academic achievement generally finding that these schools do create modest improvement in test scores (McEwan, 2000). Religious schools also expose students to different types of training, and many researchers have investigated whether religious schools reduce the incidence of disciplinary infractions, truancy, teenage pregnancy, and other nonacademic behaviors (e.g., Figlio & Ludwig, 2000).

4. Appendix Table 26.A.1 focuses on domestically funded voucher programs. A comprehensive list of currently operating privately funded scholarship programs can be found at [http://www.edreform.com](http://www.edreform.com). West (1996) contains a list of international voucher programs.

5. The average family income of students participating in the program in the early 1990s was $11,600, and the test scores of these students were below the district average as a whole and for a sample of low-income students within the district (Witte, 2000).

6. In 2002 in *Zelman v. Harris*, the U.S. Supreme Court upheld the constitutionality of the publicly funded voucher program in Cleveland.

7. It should be noted, like many choice programs, the study had a fair amount of attrition as only 27 percent of students offered the scholarship used the voucher for all years (Wolf et al., 2010, p. xxiv).


9. It should be noted that these scholarships fund only a portion of the total cost of attending private schools, and the families are expected to make up the differences. This was a conscious decision of CSF as they wanted families to have a financial stake in their decision.
One challenge these studies have had to address is attrition of the scholarship “winners” in the lottery assignments. A substantial portion of students that are randomly assigned scholarships never use the scholarships. Presumably, a portion of the scholarship “losers” would have also declined the scholarship offers. Therefore, comparing the performance of scholarship winners and losers through random assignment is complicated by not knowing who would have declined the scholarships in the group that was not assigned a scholarship. Two approaches of dealing with this challenge is to examine the effect of those offered the scholarship, whether they take the scholarship or not, to the control group (i.e., “intent to treat” analysis) or to use sophisticated statistical techniques to estimate the actual effect of attending a private school (McEwan, 2004).

It should be noted that the Charlotte study was for one year with no follow-up.

This effect was significant in all three years in New York and in the second year in Washington, though no significant effect was found in Dayton. Aggregating over the three cities, the effect of 0.28 and 0.30 standards deviations was significant in the second and third years. Greene (2000) did not break out the effects by race.

The relationship between peer effects and educational vouchers has been explored in theoretical models presented in Epple and Romano (1998) and Nechyba (2000).

The important exception is in Chile where voucher users tend to have better socioeconomic characteristics.

References


Appendix

Table 26.A.1 lists the domestically funded non-special-need voucher programs, including educational tax credit programs in place by the 2013–14 school year. As part of the table, we list the location, type of program, who is eligible, how much the subsidy is worth, and the approximate number of students served. We excluded special needs programs because almost no research has been conducted on these programs.

TABLE 26.A.1 Domestic, Publicly Funded Programs
<table>
<thead>
<tr>
<th>Location</th>
<th>Year</th>
<th>Type of program</th>
<th>Who is eligible</th>
<th>How much in the outside world</th>
<th>Appreciation number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>2000</td>
<td>Voucher</td>
<td>Children in grades K-12 in the Cleveland School District are eligible. Priority is given to families with incomes below 200 percent of the federal poverty level. Children from families with incomes above 200 percent of poverty are eligible to receive vouchers if approved by the Office of Superintendent of Public Instruction.</td>
<td>$6,000 for grades K-12 and $6,500 for high school.</td>
<td>In 2012-13 school year, 6,693 students</td>
</tr>
<tr>
<td>Colorado</td>
<td>2011</td>
<td>Voucher</td>
<td>Douglas County Pilot Program provides vouchers to any public school student who has spent at least one year in the Douglas County School District.</td>
<td>The maximum voucher value is $6,000 for grades K-12 and $6,500 for high school. Families with incomes below 200 percent of the federal poverty level receive vouchers worth 90 percent of the maximum voucher value. Families with incomes above 200 percent of poverty are eligible to receive vouchers if approved.</td>
<td>In 2011-12 school year, 494 students</td>
</tr>
</tbody>
</table>

**In 2012-13 school year, 5,365 students**

- **Students in grades K-12 in the Cleveland School District are eligible. Priority is given to families with incomes below 200 percent of the federal poverty level. Children from families with incomes above 200 percent of poverty are eligible to receive vouchers if approved by the Office of Superintendent of Public Instruction.**

- **The maximum voucher value is $6,000 for grades K-12 and $6,500 for high school. Families with incomes below 200 percent of the federal poverty level receive vouchers worth 90 percent of the maximum voucher value. Families with incomes above 200 percent of poverty are eligible to receive vouchers if approved.**

- **In 2011-12 school year, 494 students**
Emerging Alternatives to Neighborhood-Based Public Schooling

JOSHUA M. COWEN AND EUGENIA F. TOMA

Introduction

Historically, most American families have enrolled their children in public schools near their place of residence. Over the past thirty years, however, policymakers have expanded the educational alternatives available to parents (Butler et al., 2013). For example, drawing student enrollments from many neighborhoods, magnet schools were part of a larger set of policies to desegregate public education. These were followed by an expansion of intradistrict and interdistrict choice opportunities known as open enrollment. Since the last decade of the 20th century, two additional initiatives have gained momentum: charter schools and educational vouchers to be used in privately operated schools. Charter schools continue to grow in numbers and enrollments, and educational vouchers, though far smaller in enrollments, have generated considerable debate and legislation in many states. Technology has added to this choice set. In addition to brick and-mortar alternatives, virtual schools are developing at a rapid pace, both within the charter framework and as part of a home schooling option that has remained present in the American educational system since the development of public schooling.

This chapter highlights these developments. We begin with a bird’s-eye view of school choice policies by state. For each school choice option, we call particular attention to the financial and organizational incentives—codified by state legislation—that shape alternatives to the traditional sector. In broad terms, we argue that meaningful and fundamental distinctions affecting supply exist across the country. These distinctions, which cover administration, access, oversight and finance, among other domains, lead to wide variation in student access to nontraditional alternatives.

Summary of New Schooling Options

Public resources support three major forms of school choice, and public policy dictates the extent to which a fourth form, home schooling, is a viable alternative to the traditional school system. In this section, we discuss how the three general choice options—magnet/open enrollment schools, charter schools, and voucher/tax credit programs—as well as regulations for home schooling vary broadly across the United States. We organize our discussion by beginning with the choices available within the public school system itself before shifting focus to programs that successively move students away from the typically defined public sector.

Magnet and Open-Enrollment Schools

The first publicly supported school choice programs began within the traditional public school system as alternatives to the neighborhood assignment of students to schools. Magnet schools first appeared in the late 1960s, as part of the effort to desegregate schools in urban areas (Rossell, 2005). Today, magnet schools operate in nearly all states, drawing students based on interest in subjects such as science, technology, engineering, and mathematics (STEM), business, and the creative arts, or students with particular aptitude for gifted/talented
Increasingly, public school systems have expanded to provide alternatives to neighborhood-assigned schools regardless of student interest or aptitude or goals of desegregation (Schneider, Schiller, & Coleman, 1996). Such open-enrollment plans have developed in most states over the past 25 years, with Minnesota establishing the first open-enrollment framework in 1988. Open enrollment generally denotes a system in which students are free to attend any age-appropriate public school outside of their neighborhood catchment area, provided that the school has not exceeded enrollment capacity. In interdistrict systems, public funding (at least at the state-level) typically follows the student to the selected district (Carlson, Lavery, & Witte, 2011; Reback, 2008).

Some states require both interdistrict and intradistrict open enrollment. Others require districts to accept transfers from outside their jurisdictions but do not require choice within district boundaries. In still other states, districts must allow enrollment between schools, but are not required to accept students outside their boundaries. In a few states, open enrollment is voluntary or nonexistent. These policies are in place in addition to the provisions introduced by the federal No Child Left Behind Act (NCLB) that require districts to allow students to transfer within their borders to successful schools and away from schools needing improvement, as well as more specific state and district choices for students with specific physical, mental, or emotional disabilities.

State regulation of the administration of open-enrollment programs also varies (Education Commission of the States, 2013). Most states provide a mechanism for districts to opt-out of open enrollment, usually for reasons related to space or financial capacity of the district. How districts allocate space, particularly in schools with high demand, also differs. Lotteries may be used to guide oversubscribed enrollment levels, while many states prioritize students with lower levels of family income or those with special academic needs. Issues of racial isolation or segregation remain a challenge across the country, often addressed explicitly in some guidelines for maintaining racial or ethnic balance after student transfer. Some states allow districts to set such standards on their own, but provide no legislative guidance. Transportation also affects open enrollment. In many states, parents are either solely or primarily responsible for the cost of attending an out-of-district school. In some, the cost is divided between the sending and receiving districts, with school or district-level measures of academic performance further influencing these costs.

**Charter Schools**

“Charter schools” is the summary term for public schools established under state laws that waive certain regulations for schools meeting specific requirements set out by their approved application or charter. Typically this establishment rests on a contract between a particular entity and an authorizing agency—usually the state or school district. Charter schools were first authorized in Minnesota in 1991, with states such as California, Colorado, Massachusetts, Michigan, and Wisconsin following within the next two years. Although most initial charter legislation passed during the second half of the 1990s (U.S. Department of Education, 2013; Center for Education Reform, 2013), very recent authorizers include Mississippi (2010), Maine, and Washington (2012). The first column in Table 27.1 summarizes charter availability nationwide.

Table 27.1 Distribution of Charters, Vouchers, and Home School Policies by State
Depending on the state, charters may be "start-up" schools—new schools open for the first time—or conversion schools reorganized under a charter from existing traditional public school or a private school. In some cases, conversion from a traditional public school simply represents the establishment of a new, separately administered school that is still directly responsible to the original school district. Although the vast majority of charter schools nationwide (85 percent in 2010–11) are run by nonprofit organizations or school districts, often with philanthropic support, nearly one in 10 charters are operated by for-profit management companies called Education Management Organizations, or EMOs. A small number of virtual charter schools providing online education are also operating in several states, sometimes as part of an EMO (National
**Vouchers and Tuition Tax Credits**

In recent years, states and other jurisdictions have turned increasingly to the private sector to provide alternatives to traditional public schooling. Although private schools overall have experienced declining enrollments during the past decade and a half (Aud et al., 2012), the use of public resources to support private school attendance—especially for historically disadvantaged students—has grown. Whether via vouchers—direct public subsidies for private tuition—or deductions or credits administered through the tax system, these programs are based on the notion that a market for educational services should include public subsidies for private competitors to maximize consumer choice (Friedman, 1962). They are also consistent with the expansion in recent decades of more general contracting between public service organizations and private providers (Carlson, Cowen, & Fleming, 2013).

The first large-scale, publicly supported voucher program began in 1990 in Milwaukee, Wisconsin, although Maine and Vermont have offered town-tuition programs for students in districts without public schools since the late 1800s. Columns 2–4 of Table 27.1 display the availability of voucher or tax-credit or deduction programs nationwide. As with other and more widely spread forms of choice, however, the presence of public support for private tuition may overstate its availability within a particular state because voucher programs are often subject to various restrictions. Restrictions may be geographic—e.g., programs operated only for residents of a large urban jurisdiction or its contiguous school district—or they may be based on income-eligibility limits. Some programs, notably Florida’s McKay Scholarship program, have historically been targeted to children with special academic needs. In other states, such as Ohio and Louisiana, vouchers are tied at least partially to low academic performance of a student’s assigned public schools (National Conference of State Legislatures, 2013).

**Home Schooling**

In one sense, home schools represent the oldest form of school choice, as family-based education predates even the common school in the history of American education. As we elaborate below, home schooling represents a particular aspect of choice in which the suppliers are the same as the consumers. Compulsory school attendance laws raised the legality of home schooling in public debate until a Supreme Court case (Pierce v. Society of Sisters, 1925) clarified the parental right to meet educational requirements for their children by supplying a home school option. Today, the matter is largely one of regulation, and state laws governing home schooling vary as broadly as those authorizing charters and vouchers. In column 5 of Table 27.1, we have organized these regulations into categories of “high,” “low,” “moderate” and “none.” This range extends from laws that require home-schooled students to take state tests and require approval of home-based curricula (high); to laws that stipulate nothing but that parents must notify their districts of their intention to home school (low), to few or no regulation at all (none). Although most states have some mix of these requirements—and many indeed have none—few include major restrictions. The strictest laws are found in New England and the mid-Atlantic, including Massachusetts, where much of the early push for public schooling occurred in the mid-19th century.

**Current Landscape: Variation in Voucher, Charter, and Home School Educational Suppliers**

In this section, we consider in more detail the three main alternatives included in Table 27.1. These three options differ from magnet and open-enrollment programs in that students are not in traditionally organized public schools, whether neighborhood-assigned or otherwise.
Charter Schools

One of the distinguishing features of charter schools relative to the traditional public schools is the mechanism for enforcing the accountability standard. If a school is unable to perform according to its charter or if a school is not financially viable, the school is subject to closure. About 15 percent of charter schools have closed over their relatively young history (Center for Education Reform, 2011). The reasons for closure are typically financial problems or mismanagement rather than direct academic concerns, and typically are most problematic for schools in their early days of operation. Evidence suggests that the charters become more stable and perhaps more productive as they age (Carruthers, 2012; see also Bifulco & Bulkley, this volume).

Since Minnesota adopted the nation’s first charter school law in 1991, charter schools have expanded across the country. Although growth rates were initially slow, the pace began to increase in the early 2000s. Table 27.2 displays the number of charter schools by state over four time periods since the mid-1990s. Charters, or their authorizing legislation, are currently present in all but eight states, including the District of Columbia. Within the charter states, more than 6,000 charter schools are currently open and serving two million students, which is approximately 4 percent of the nation’s 48 million school-aged children (Consoletti, 2012). State authorization policies differ across a variety of dimensions, including teacher certification and collective bargaining requirements, curricular independence, performance reporting, and control over school finances. Two components of each state’s charter school law directly dictate charter availability: the extent to which states cap either enrollment and/or, the number of schools; and the extent to which states limit which entities can initiate a charter.

California, Arizona, Florida, and Wisconsin contain over 40 percent of all charter schools. There are also large concentrations in New York, Michigan, and Louisiana (National Alliance for Public Charter Schools, 2012). Within states, larger concentrations of students are enrolled in charters in urban areas, because rural areas are less likely to provide the enrollment base needed to support charter schools. As a result, the supply of charters in rural areas is very small. Thus, discussions of charter schools typically refer to an alternative type of schooling for children in urban or suburban areas.

Financing Charter School Attendance. Students who enroll in charter schools pay no fees or tuition; rather charters receive public funding on a per-pupil basis that they may supplement with grants or private philanthropic funds. When a student moves from a traditional public school to a charter school, the public funding that would have gone to the traditional school for that student typically follows the child to the charter school. The specifics of the funding formula often depend on the state law and the authorizer of the charter school (Nelson, Muir, & Drown, 2000, p. 30). For the majority of charter schools authorized by a district, charters receive the district’s average per-pupil revenues, average per-pupil expenditures, or funds determined by an alternative per-pupil formula. In general, when a state entity authorizes a charter school, state formulas for per-pupil funding follow one of two methods. First, the state may use the district’s formula to establish funding levels. This method allows the state agency to account for the grade level, special needs, and low-income status of the student. Alternatively, the state may transfer funding based on a school district’s average per-pupil funding level. In this method, the state assumes that the charter school will enroll a representative sample of the district’s students (Consoletti, 2012).

Table 27.2 Number of Charter Schools in Operation by State

Table 27.2 Number of Charter Schools in Operation by State

407
It is possible that these resources come at a net cost to traditional public school districts. How states or districts calculate (or divert revenues for) the per-pupil funds received by charter schools determines the fiscal impacts. These impacts depend in part on the extent to which students attending charters would have attended a public school in the districts or would have enrolled in a private school (similar hypotheticals determine fiscal impacts of other choice programs; see, e.g., Costrell, 2009). In New York State, for example, Bifulco and Reback (2014) find that charter school enrollments have negative fiscal consequences for the Albany and Buffalo school districts, especially under scenarios in which the districts are unable to correspondingly reduce spending proportionate to their loss of revenue. Cordes (2013) finds, however, that, once established in a district, charter schools in New York City lead to average increases in total, general education, and instructional spending per pupil. Nationwide, preliminary evidence shows that charter schools located in a district spur growth in overall per-pupil revenues of that district overall. Such a pattern could occur if charters increase overall spending by attracting students who previously attended private schools (Buddin, 2012) or by increasing the overall support for public education with new alternatives that ultimately lead to increased revenues for public education in general (Jones, 2014). Whether such revenues represent positive or negative developments from the perspective of the traditional public schools remains uncertain and is an emerging line of research in public school finance.

Educational Management and Charter Management Organizations. The organizational structure of charter schooling differs widely across the country. In general, charter schools may belong to a network of providers or may be free-standing individual schools. For-profit education management organizations (EMOs) are permitted to contract with public schools to oversee operations in at least 35 states and have become one of the more controversial aspects of the charter movement (Miron & Gulosino, 2013). Nonprofit charter management organizations (CMOs), such as the Knowledge Is Power Program (KIPP) or Aspire, operate schools according to shared educational missions, often with philanthropic support (Emerson, 2013). Freestanding, individually organized and administered charter schools still comprise the majority of providers nationwide, although that is not the case in all states. In Michigan, for example, the majority of charters are held by EMOs, with very few CMOs. In Florida, the majority of charter schools are freestanding, reflecting the national pattern, while in Texas, CMO charters have grown steadily since 2008, becoming a majority of providers by 2010.

<table>
<thead>
<tr>
<th>State</th>
<th>2007-08</th>
<th>2008-09</th>
<th>2009-10</th>
<th>2010-11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Utah</td>
<td>1</td>
<td>23</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>Colorado</td>
<td>26</td>
<td>130</td>
<td>100</td>
<td>150</td>
</tr>
<tr>
<td>Nebraska</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>California</td>
<td>81</td>
<td>781</td>
<td>1,060</td>
<td>1,060</td>
</tr>
<tr>
<td>Idaho</td>
<td>12</td>
<td>199</td>
<td>89</td>
<td>150</td>
</tr>
<tr>
<td>Montana</td>
<td>0</td>
<td>15</td>
<td>21</td>
<td>130</td>
</tr>
<tr>
<td>Illinois</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>16</td>
<td>35</td>
<td>13</td>
<td>35</td>
</tr>
<tr>
<td>Michigan</td>
<td>16</td>
<td>35</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Minnesota</td>
<td>13</td>
<td>27</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>Missouri</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>New York</td>
<td>19</td>
<td>25</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>North Dakota</td>
<td>14</td>
<td>16</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>Ohio</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>13</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Oregon</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>0</td>
<td>82</td>
<td>135</td>
<td>135</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>0</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>South Carolina</td>
<td>0</td>
<td>20</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>South Dakota</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tennessee</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Texas</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Utah</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vermont</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Virginia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Washington</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Wyoming</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Sources: NCES, National Center for Education Statistics; Education Week; Education Week; and Education Week.
The development of EMOs and CMOs since the mid-to late 1990s helps explain the concurrent growth in charter schooling (Miron & Gulosino, 2013). Legislation authorizing charter schools may, depending on the state, provide a ready point of entry into the market for entrepreneurial vendors. At the same time, the services provided by EMOs and CMOs may alleviate some of the difficulties faced by charter school organizers as they establish their school. Finding and maintaining a school building, obtaining overhead and ongoing funding, and developing curricula are all tasks for which founders may plausibly turn to an EMO or CMO for support (Miron, 2008). In addition, under No Child Left Behind, conversion to a charter school and/or delegation to an EMO are options specified for failing schools (Hess & Petrilli, 2006; Zimmer et al., 2007). Some observers refer to CMOs simply as "nonprofit EMOs," (e.g., Miron, 2008), but there are distinctions beyond profit motivation. Although the prevalence of EMOs has grown steadily over time, CMOs have experienced rapid growth in very recent years (Bifulco & Bulkley, this volume; Farrell, Wohlstetter, & Smith, 2012). The explosion of CMOs is explained partly by funding from larger nonprofit foundations and federal encouragement but also by the promise of replicating perceived success of certain CMO management models across networks of schools (Farrell, Wohlstetter, & Smith, 2012; Furgeson et al., 2012). Such common missions, and the affiliate network, define the CMO structure. The CMO model relies on collaboration between schools, a balance between individual school autonomy and collective capacity across the network, and the linking of each school’s practice to the successful reapplication of the CMO’s educational brand that can presumably be distributed quickly and without dilution in multiple settings (Farrell, Wohlstetter, & Smith, 2012), often where individual start-ups fail to gain traction. In some areas, local political support encourages these charter networks (Bifulco & Bulkley, this volume), which are clustered in several states that permit relative independence from the public school system as well as in urban areas (Furgeson et al., 2012). In addition to the particular sense of mission that may unify schools across the CMOs, CMO models themselves may have common elements. A recent example is the “no excuses” model that emphasizes student outcomes, additional instructional exposure, and teacher hiring and retention. This model characterizes KIPP and others, such as Achievement First and Harlem Children’s Zone in New York City (Bifulco & Bulkley, this volume; Angrist, Pathak, & Walters, 2011; Dobbie & Fryer, 2011).

How Elastic Is the Charter School Supply? What explains the rapid expansion of charter schooling over the past decade? One way to frame the question is by asking, "Will the supply of places in charter schools expand so long as there is demand for them?" (Hoxby, 2006, p. 15). A priori, we might expect charters to be more responsive to consumer demand in those states where the laws are relatively conducive to entering the market as a new school or, for existing schools, where enrollment is expanding. Some early evidence is consistent with this expectation. Hoxby (2006) estimated the elasticity of charter school supply by examining the date on which a state passed charter school laws, the expected costs associated with certain elements of the laws, such as the amount of public funding available to the charters relative to that available to traditional public schools, and the ease of authorizing charters. Overall, Hoxby found that the supply of charter schools is "greater in states whose laws create an even playing field between charter schools and regular public schools" (2006, p. 16). Some of the main elements of an even playing field included fiscal autonomy, operational autonomy at start-up, and per-pupil funding levels commensurate with the traditional public schools. She also found that the number of charters is more responsive to the nature of the legislation at the elementary and middle school levels than at the high school level. As noted later in this chapter, the same appears true for home schooling, which suggests that the decisions of suppliers at the high school level differ distinctly from those of suppliers of schools for younger children. This pattern may reflect the higher costs of entering the high school market.

Within a given state, charter schools do not locate randomly. Like business owners, charter operators have an incentive to open where potential demand is high and competition from rival firms is low. Charter organizers may be attracted to districts where the current public option, for one or many reasons, does not satisfy the education preferences of some households living within districts’ boundaries. These organizers, for example, may see poorly performing districts as an opportunity to provide a new education option for dissatisfied residents. Socioeconomically and racially diverse districts, while perhaps satisfying certain portions of the population, are more likely to leave other segments dissatisfied than more homogeneous districts that have fewer conflicting demands. Previous research supports this conclusion as charters locate where populations are diverse in terms of race, income, and adult education levels (Glomm, Harris, & Lo, 2005;
Stoddard & Corcoran, 2007). Additionally, recent work on charter location decisions found that within districts, charter schools tend to locate near concentrations of low-income and minority students (Bifulco & Buerger, 2012), presumably because these students are least satisfied with the traditional public schools.

In summary, the empirical evidence supports the view that legal and financial variables that define the state’s charter climate help to explain the location and growth in the number of charter schools. Demographic characteristics of the population also play a significant role.

Vouchers and Tuition Tax Credits

The use of public resources to support private education is perhaps the most controversial aspect of school choice policy. Historically, the voucher debate centered around the constitutional question of whether government funding for private, usually religious, schools violated the First Amendment’s prohibition of established religion. In 2002, however, the U.S. Supreme Court ruled that vouchers are constitutional in programs whose primary objectives are educational rather than religious; where public funds are not general subsidies for particular religious schools; for which eligibility is extended to a wide spectrum of students; which are demonstrably neutral with respect to religion; and, crucially, where a “genuine choice” between religious and secular private schools is available (Zelman v. Simmons-Harris, 2002). Many states, however, have more restrictive constitutional language. Thus, while vouchers have cleared the federal constitutional hurdle, there are lingering disagreements over the use of public funds devoted to education in religious institutions.

The first statewide program began in Florida in 1999 as the McKay Scholarship for students with disabilities. The state’s more general A+ Opportunity Scholarship Program, which began at the same time, was declared unconstitutional by the Florida Supreme Court in 2006 on the grounds that a voucher program violated the state’s commitment to a “uniform” system of public education (Richard, 2006). Under McKay, new vouchers are available to any student with a documented Individual Education Plan (IEP) enrolled in a traditional public school during the previous academic year. The size of the voucher is limited to the equivalent of funds the student’s previous public school received (Florida Department of Education, 2013). Similar plans for IEP students are available in states such as Georgia, Oklahoma, Utah, and, in addition to the other vouchers available more broadly, Ohio.

Both Ohio (in 2005) and Wisconsin (in 2011 and 2013) have recently expanded voucher programs outside Cleveland and Milwaukee, respectively. In Ohio, students must have attended low-performing public schools to be eligible, and a statewide cap remains at 60,000 students (National Conference of State Legislatures, 2013). In Wisconsin, the 2011 expansion essentially created a Milwaukee-style program in the nearby city of Racine (2011 Wisconsin Act 32), while the 2013 expansion provides vouchers to parents on the basis of availability in a small number of schools across the state. Ohio and Wisconsin are joined by Indiana, Louisiana, and Washington, DC, to offer direct voucher programs provided for students without special needs (Workman, 2012).

In Indiana, Washington, DC, and Wisconsin, eligibility is still linked to means tests, albeit at limits above the federal poverty rate. In Ohio and Louisiana, the voucher is tied to the state standards-based accountability. In these states, students in low-performing public schools are eligible to transfer to a select private school. School outcomes are measured according to the states’ respective accountability schemes that summarize a variety of indicators to form final school ratings. In Louisiana, low performance is defined broadly, including any school ranked C or below on an A–F grading system. In Ohio, a variety of thresholds can apply, including any student in the lowest 10 percent of schools rated (Workman, 2012; Ohio Revised Code 3301–11–03).

Apart from programs that directly subsidize private tuition, several more states use the tax code to provide access to private education for their residents. The systems that most closely resemble school vouchers are those that offer scholarships, such as vouchers from a particular fund into which either corporations or individual residents can contribute voluntarily (Welner, 2008). The state support for this fund comes from the tax incentive to donate: state tax liability for donors is reduced according to the size of the contribution and is typically subject to calendar-year limits on dollar amount. The scholarships distributed from the state fund otherwise closely resemble vouchers in their eligibility requirements for students and families (typically tied to income tests or special academic needs) but are also capped at the state level by total program expenditure

410
rather than enrollment (Workman, 2012). In addition to these direct scholarships, several more states offer individuals either tax credits or tax deductions for the cost of private school tuition (see Table 27.1).

Although state-specific legal challenges remain for certain features of individual program designs, the national debate over public support for private schools has largely shifted since Zelman from legal to policy terms, many of which directly pertain to provision and supply of this alternative to traditional public schooling. Whether vouchers or tax credits are disproportionately used by students with higher academic or intellectual capabilities, whether they are more efficient providers of education than traditional public schools, or whether publicly supported private schools apply competitive pressure to traditional public schools are all central elements of the debate. A new dimension has developed over testing and accountability requirements facing private schools that accept voucher programs.

Student Demand. Means tests in most program designs have limited the extent to which voucher programs make private education available to students who are comparably better off in terms of family income or other resources. In addition, the targeting of many programs toward students with special academic needs further restricts those vouchers away from academically advantaged individuals. As a consequence, studies that have considered differences between students who apply or make use of vouchers and those who do not have found that the programs generally serve students they were intended to reach. In the most recent evaluation of Washington, DC’s voucher program, applicants performed similarly to nonapplicants on prior tests scores and were more likely to be enrolled in special education before applying (Wolf et al., 2005). Similarly, recent evidence from Florida’s tax scholarship program indicates that the lowest performers and those from lower-performing public schools sought the newly formed scholarships (Figlio, Hart, & Metzger, 2010; see also Zimmer & Bettinger (this volume) for greater detail). It is important to note that some families may seek private schooling for nonacademic reasons. The majority of private schools remain religiously based (Aud et al., 2012), and parents of voucher recipients with otherwise similar demographic characteristics may seek private schools precisely because of the religious element (Fleming et al., in press).

Academic Outcomes. Whether vouchers improve student achievement remains a major point of contention and is the focus of more extended discussion elsewhere in this volume (Zimmer & Bettinger). Some studies (Figlio, 2011; Howell et al., 2006; Rouse, 1998; Greene, Peterson, & Du, 1998) have indicated positive gains, depending on student grade, race, subject matter, or program site. Others have found little difference associated with voucher attendance (e.g., Witte, 2000; Plucker et al., 2006; Metcalf et al., 2003), and the most recent Washington, DC, evaluation indicated academic growth in only one year out of four for the entire sample (Wolf et al., 2013). However, the DC evaluation also found significant differences in the probability of graduating high school, as has recent reanalysis of the privately funded voucher program in New York City from the late 1990s (Chingos & Peterson, 2013).

Competitive Effects. The rapid expansion of voucher and voucher-like tax credit programs also brings the question of competitive effects on public schools—improvements made in the public system under threat of voucher competition for students—to the forefront. In addition to increased availability to alternative providers in real terms, such expansion may heighten the perception among public school administrators that improved performance in that sector is required (Barrow & Rouse, 2008). Early evidence from Milwaukee suggested that public schools do respond positively to the presence of voucher schools (Hoxby, 2003; Chakrabarti, 2008), although it is not clear whether these gains are in fact a response to competitive pressures (Carnoy et al., 2007). A larger body of evidence from a now-defunct voucher program in Florida also supports the notion of competitive threats, indicating positive responses among low-performing public schools facing the loss of students to the private sector (Rouse et al., 2007; Chiang, 2009; Figlio & Rouse, 2006; West & Peterson, 2006; see also Gill & Booker, this volume, for greater detail).

Accountability. With the exception of programs targeted specifically to special needs students, all of the laws providing voucher programs to students on the basis of need also require their receiving private schools to administer a standardized testing regime (Workman, 2012). In most cases, private schools are explicitly required to test voucher students using the state’s own accountability testing instrument. At the same time such requirements often do not subject the receiving schools to the public reporting and accountability regulations faced by traditional public schools for their students. Voucher advocates argue that market forces are sufficient to ensure quality private schooling. For example, the pro-voucher Friedman Foundation argues that “what gives the concept of accountability real teeth is a parent’s ability to choose a school freely”
(Friedman Foundation, 2013). Recent empirical evidence from Milwaukee has shown that the test scores of students in voucher schools improved considerably after the state legislature began to require public release of school-level test scores, suggesting that the market alone did not maximize private school performance (Witte et al., in press; Carlson, Cowen, & Fleming, 2013). Regardless, accountability represents perhaps the next point for debate on vouchers, as even advocates for school choice are clearly divided over the extent to which private schools receiving voucher funds should face the same accountability requirements as their public counterparts (Finn et al., 2009, p. 3).

**Home Schooling**

Although home schooling is viewed by many as a modern form of school choice, its roots can be traced to the earliest settlements within the United States (Houston, 2000). Parents tutored their children or paired with other parents to hire tutors throughout the 17th and 18th centuries (Cremin, 1976). Home schooling continued to flourish late into the 19th century and some have linked its continued popularity, as well as the rise of more traditional private schooling, to religious conflicts within communities (Lines, 1986). For example, community-financed public schools began to replace the private tutor in the latter part of the 19th century. These community schools were accepted by the townspeople, and instruction in religion was a common part of the curriculum either explicitly or implicitly. This arrangement worked as long as members of the population viewed religion in fundamentally similar ways. But the 19th century welcomed a new wave of immigration to the United States—this one dominated by Catholics. Some of the more popular textbooks, such as the New England Primer, were openly anti-Catholic. Philadelphia was the site of a riot in 1844 over whether the Protestant or Catholic version of the Bible would be used in the community-financed schools. This controversy led to a call by the Catholic Council urging the education of Catholic children in their own, Catholic-based schools (Lines, 1986).

An issue into the 20th century was whether parents had to enroll their children in any type of schooling. Compulsory attendance laws passed by all states (Mississippi was last) required parents to choose some form of schooling for their children—no schooling was not an option. Oregon went further and attempted to require that all students attend public schools. It did so by writing compulsory attendance laws in such a way that all other forms of schools, private or home, were in defiance of the laws. The U.S. Supreme Court ruling *Pierce v. Society of Sisters* in 1925 overturned the Oregon law. While state laws in the aftermath allowed alternatives—private or home schools—to the public schools to satisfy compulsory attendance, laws regarding home schools were not always well specified, and for much of the 20th century the home-school population remained quite small and did not constitute a significant form of school choice.

In one sense, a rebirth of home schooling began in the 1970s with another Supreme Court decision, *Wisconsin v. Yoder*. In 1972, the ruling stated that the “state cannot compel a child to attend public school in the face of strong religious objections and when the state’s interests in the education of the child is adequately served by an alternative program” (Lines, 1982, p. 121). Political groups as diverse as organized religion and the American Civil Liberties Union began to support the parents’ rights to home school. Today, states almost uniformly accept the underlying premise that the constitutional responsibility to provide for education does not extend to controlling the means by which it is delivered.

All states now allow some form of home-based education. The degree to which the general laws governing home schools are restrictive or permissive are included above in Table 27.1. Thirty-five states have laws that provide specific rules for home schools. Some states merely request that parents register with a particular agency so that compliance with compulsory attendance laws can be monitored. A significant majority of states place no restrictions on parental qualifications for educating their children at home. At least one state, Idaho, allows dual enrollment in public and home schools, and California provides a $1,000 voucher for home-schooled students to purchase district-approved textbooks. There appears to have been a return to the cultural as well as legal acceptance of home schooling as a form of school choice.

Home schools represent a unique way of thinking about supply alternatives to traditional public schools. The households that demand home schools are actively or passively supplying the schooling. Because the states count numbers of students with differing degrees of precision, it is somewhat challenging to estimate home school enrollments. Nevertheless, the National Center for Education Statistics estimates that the percentage of
students who were home schooled in 1999 was approximately 1.7 percent (NCES, 2009). By 2007, the percentage had grown to an estimated 2.9 percent of the school-age population, and in 2012, the estimated percentage was 3.0 of the students enrolled. While the growth rate seemed large in the early 2000s, enrollments seem to have held relatively stable more recently.

Although it has been legal to home school in all states since 1993 (Somerville, 2005, p. 137), analysis across states has been limited largely because data are not strictly comparable across states, and some parents may be less than fully compliant with surveys on their home schooling practices. Even so, raw numbers suggest significant variation across states in this type of schooling option. In a few states, including Wyoming, Utah, South Dakota, and Vermont, less than 1 percent of students appear to be home schooled. At the other extreme, in South Carolina, Tennessee, Virginia, and Texas, more than 10 percent of their school-age population is reportedly home schooled. Note that this variation may be attributable, in part, to the way in which the states count home schools. Tennessee, for example, allows affiliation with church schools to count as home schools, so numbers are not strictly comparable to home schools across other states.

There is little evidence that specific state regulations have played a role in the growth rate of home schools, although their legal recognition has certainly contributed to the overall upward trend in households choosing home schools. Technological advances such as home computers and the Internet almost certainly contributed to the growth of home schools in the 1990s and early 2000s. These technological changes reduced the cost of networking and acquiring the materials required by the home school household or group (see next subsection).

Rather than examining the supply side of home schooling as a function of state laws, most research has been based on a variety of sample and/or survey data and focused on the household as the unit of decision making about the choice to home school. From this research, we have learned a few things about the demand and, thus, the supply of home schools. First, two-parent households are more likely to supply school at home than households headed by single parents or adults (Houston & Toma, 2003; Isenberg, 2007). This is not surprising given the nature of the supply decision as homeschooling tends to be more labor intensive for the household than any other school choice decision. Indeed, Howell and Sheran (2008) and Bauman (2001) found that households that supply home schooling are more likely to have only one of their adults in the official labor market.

The disparity in income of the population of a school district also appears to influence the use of home schooling. Presumably families within a school district with a homogeneous population are likely to be more satisfied with the traditional public schools and find less value in creating alternative forms of schools of any type including home schools than districts with large income disparities. Evidence from both districts and a sample of states supports this thesis (Houston & Toma, 2003). In a similar vein but using household data, Isenberg (2007) found the use of home schools to be greater in areas where the quality of public schools is low and private availability is low.

Self-reported reasons for home school supply generally support the few studies based on either administrative or survey data. At different points in time, households have reported that they educate children at home because they are dissatisfied with the environment or the quality of the public schools or, more specifically, that the traditional public schools fail to provide religious instruction. But more generally, the link between religion and home school supply remains unclear. Belfield (2002) noted that fewer than half of the home school students who participated in the SAT declared a religious affiliation. Using household survey data, in contrast, Isenberg (2007) found that a child from a “very religious household” was significantly more likely to be home schooled than other children, controlling for other variables.

In summary, the amount of home schooling cannot be estimated with precision because the data are not of sufficient quality. But even beyond the evidence, conceptually, it is reasonable to argue that home schools are supplied because the alternatives do not provide the households with the same degree of satisfaction that can be achieved through the home. Whether dissatisfaction is attributable to the lack of religious instruction in public schools, the academic quality of public schools, or concerns regarding drugs or safety at schools cannot be answered with the scientific rigor that education policy researchers would typically require. Instead, we can recognize that home schooling does now constitute the school type of choice for roughly 3 percent of the school-age population. Future research on these questions is warranted.

A Note on Cyber Schools: The Intersection of Charter and Homeschool Supply

413
Virtual and homeschool charters—often referred to as “cyber schools,” “nonclassroom-based charter schools,” “cyber charter schools,” and “e-schools”—are a small but growing part of the state charter school movement, as well as important institution vehicles for implementing digital and distance/online learning programs (Huerta, d’Entremont, Gonzalez, 2009). These cyber schools are public schools that function under the same state rules and regulations as traditional charter schools although they generally deliver 100 percent of their courses online.

According to the National Alliance for Public Charter Schools, more than 30 states now enroll students in virtual charter schools (Ziebarth, 2014), while more than 310,000 students are enrolled in these schools, accounting for close to 1 percent of student enrollment nationwide (Griffith, 2014). Furthermore, in the 2012–13 school year, student enrollment in virtual charter schools grew by 20,000 students (7.5 percent) over the 2011–12 school year (Griffith, 2014). Both Pennsylvania and Arizona in particular have seen significant growth in their virtual charter school systems. In Pennsylvania, one virtual charter school enrolls almost three times the number of students as the average district in the state, while Arizona leads all states in virtual charter school enrollment with 3.9 percent of their student population in virtual charters (Griffith, 2014).

These schools have unique characteristics, such as unlimited school size and enrollment borders, and traditional funding formulas may be inapplicable (Griffith, 2014). Case study analysis from Pennsylvania and California indicates that more consistent and transparent accountability mechanisms and clearer enrollment boundaries must be specified in policy language in order to mitigate serious issues related to educational quality and the mismanagement of public funds by third parties whose primary focus is profitability and not student learning (Huerta, d’Entremont, & Gonzalez, 2009).

Owing to the small number of existing cyber schools, data on their educational impacts are relatively scarce, although some preliminary analysis suggests that policy-makers should approach this form of schooling with caution. In an analysis of California data, Buddin and Zimmer (2005) compare classroom-based charters and cyber charters, finding that cyber charter school students performed considerably worse than other charter students as well as public school students. A 2001 report by KPMG praised cyber charters for their “innovative approaches,” but was skeptical of their “long-term viability” (KPMG Consulting, 2001, p. 5). State and local education leaders are concerned that virtual charter schools may not be able to adequately monitor student attendance, teaching and learning practices, or the misuse of online programs by nonenrolled students (Herold & Cavanaugh, 2014). As Huerta, d’Entremont, and Gonzalez (2009) note, both cyber charter models have attracted large numbers of formerly homeschooled students, remaining an important trend to monitor as states and legislators consider this model of schooling, particularly given the funding challenges created if there is an initial influx of homeschool students into these programs.

Discussion: What Would Equilibrium Look Like for the Supply Side in the Education Market?

In recent years, a large number of alternatives to traditional, neighborhood public schools have emerged (Butler et al., 2013). While charter schools and educational vouchers are the most visible features of this movement, less visible features include the expansion of home schooling and virtual schooling. At present, there are more educational options available to American citizens, at either direct or indirect taxpayer expense, than ever before, with some parents continuing to opt for self-supplied home schooling for their children as well.

Whether students who learn in alternative schools realize gains in their educational achievement relative to their peers in traditional public schools is beyond the scope of this chapter. But economic theory predicts that such a system in which families may select from a variety of options—neighborhood schools, schools in other districts, magnets, charters—or have access to options outside public schools should maximize productivity for each provider. As this chapter shows, whether this potential has been or may be realized depends to a large degree on state-specific features influenced by laws and regulations. A more accurate depiction may, in fact, highlight a variety of state and substate markets for educational services, some in which the vast majority of students have access to at least one public or nonpublic alternative to traditional neighborhood schools and
others where only a small number of students have meaningful choices. A host of financial and organizational features of different choice programs varies across the country, typically determining which alternatives are available to individual students.

Although we have generally described this variation in terms of state statute, policy and regulation, and characteristics of the population that drive demand, differences in the supply of alternatives occur along a variety of other dimensions as well. We have discussed only tangentially other factors, such as asymmetries in information about the existence of these alternatives, eligibility requirements, and the academic quality of available schooling options, each of which is related to the standard idea of schooling as a means for acquiring knowledge and skill. Left almost entirely unexplored is the availability of schools that meet different geographic, cultural, and historical definitions of education itself.

Twenty years or even a decade ago, an essay on the supply of alternatives to the public sector would have necessarily employed a speculative tone. Vouchers existed only in their pilot stage, charters were largely boutique providers, and home schooling was an idiosyncratic feature of the educational landscape. If the vast majority of American children still learn in their neighborhood public schools, a rapidly growing number have a variety of other academic options. We need not prognosticate, nor advocate, to note that these developments show little sign of abating.

Acknowledgments

We are grateful to excellent research assistance by Sarah Galey, Peter Jones, and Danny Woodbury.

Notes

3. Because the estimated model specified the natural log of charter enrollment, and because most of the explanatory variables were operationalized either as natural logs or were inherently percentages, the estimated relationships are interpretable as elasticities.
4. In Colorado, Douglas County School District has authorized a local voucher program called the Choice Scholarship Pilot Program, the implementation of which is pending court appeals. See https://www.dcsdk12.org/legal-counsel/choice-scholarship-program.
5. Although most have been subject to official or quasi-official program evaluations, e.g. Wolf et al. (2013); Witte et al. (2008); Metcalf et al. (2003).

References


Wolf, P. J., Kisida, B., Guttmann, B., Puma, M., Eissa, N., & Rizzo, L. (2013). School vouchers and student...


Section VI
Race, SES, and Achievement Gaps

SECTION EDITOR: SUSANNA LOEB
Introduction

Racial, ethnic, and socioeconomic disparities in academic achievement remain a stubborn feature of U.S. schooling. National studies consistently show that the average Hispanic student and non-Hispanic black student scores well below the average non-Hispanic white student on standardized tests of math and reading skills. Likewise, the average student from a low-income family scores much lower on such tests than students from higher-income families. Considerable attention has been focused on achievement gaps, particularly the black-white achievement gap. Scholars and educators have suggested a number of possible explanations for the gaps, and policymakers, principals, and teachers have tried a range of remedies. As this chapter documents, however, the gaps persist despite these efforts. Moreover, our understanding of the causes and patterns of these achievement gaps is far from complete.

How Have Achievement Gaps Changed Over Time?

The best source of data on how achievement gaps have changed over time comes from the National Assessment of Educational Progress (NAEP).¹ NAEP includes two different assessments of the math and reading skills of nationally representative samples of students.² The first of these—Long-Term Trend NAEP (NAEP-LTT)—is given every two to four years to a nationally representative sample of children aged nine, 13, and 17. Because the tests used for NAEP-LTT have remained essentially unchanged since their first administration in the early 1970s, they provide a consistent instrument to evaluate achievement trends (see for example, Grissmer et al., 1998; Hedges & Nowell, 1998; Neal, 2005).

The second of the NAEP assessments has been administered roughly every two years since 1990 and is sometimes referred to as "Main NAEP." The content of the Main NAEP tests is updated periodically to reflect changing curricula, so that it is more appropriate than NAEP-LTT for providing information on how students perform on tests of the material taught in their schools. While this feature has obvious appeal for investigating current achievement patterns, the changing nature of the assessments may complicate comparisons of achievement levels across years. Trends in achievement gaps, however, are less sensitive to small changes in test content than are trends in achievement levels, so Main NAEP is also useful for assessing trends in achievement gaps. Using either test, researchers typically report gaps in standard deviation units.

One additional difference between the two NAEP datasets is worth noting. NAEP-LTT is administered to age-cohorts (nine-, 13-, and 17-year-olds), while Main NAEP is administered to grade-cohorts (fourth, eighth, and 12th graders). Because the populations sampled for the two assessments differ from one another, and because patterns of grade retention differ among racial, ethnic, and socioeconomic groups and may differ over time, the gaps and their trends measured by NAEP-LTT and Main NAEP would likely differ from one another even if the same tests were used in each. This will complicate interpretation of trends in measured gaps, since the average grade level of each age-cohort and the age composition of each grade-cohort may differ across population groups and over time.
In Tables 28.1 and 28.2, we report trends in NAEP achievement gaps. Table 28.1 reports 1971–2012 trends in black-white and Hispanic-white math and reading gaps and gaps between students whose parents have only a high school diploma and those with at least a four-year college degree, as measured in standard deviation units on the NAEP-LTT tests. Table 28.2 reports 1990–2013 trends in black-white, Hispanic-white, Mexican-white, and Asian-white math and reading gaps, measured in standard deviation units on the Main NAEP tests.

Trends in Black-White Achievement Gaps

Trend data on black-white achievement gaps are relatively clear: the black-white achievement gaps in both math and reading narrowed from the early 1970s through the late 1980s (Grissmer et al., 1998; Hedges & Nowell, 1998, 1999; Neal, 2005). This pattern is evident not only in NAEP, but also in SAT score trends (Ferguson, 1998) and in comparing nationally representative samples of students in other large scale studies from the 1960s through 1992 (Hedges & Nowell, 1999). The black-white gap widened in the early 1990s (Ferguson, 1998; Neal, 2005), but NAEP-LTT results show that the black-white gap has been narrowing consistently since 1999 (see Table 28.1). Likewise, the Main NAEP results also indicate a narrowing of the black-white math and reading gaps in fourth and eighth grade since the early 1990s (see Table 28.2).

A wide variety of explanations have been proposed for the trends in the black-white test score gaps, though none are conclusive. Grissmer, Flanagan, and Williamson (1998) argue that trends in black and white students’ family characteristics can account for only a small portion of the narrowing of the black-white gap through the late 1980s; they argue instead that school policies (including desegregation) account for much of the reduction in the gap. Among other things, they demonstrate that achievement gains by region were greatest for blacks in the southeast, which experienced desegregation, and lowest in the northeast, which experienced increased segregation. Other research also suggests that segregation patterns and trends are strongly associated with racial disparities in educational outcomes, including both achievement and attainment (Card & Rothstein, 2006; Guryan, 2004; Johnson, 2011; Vigdor & Ludwig, 2007). Most recently, Condron, Tope,
<table>
<thead>
<tr>
<th>Test year</th>
<th>Black-white gap Reading</th>
<th>Black-white gap Math</th>
<th>Hispanic-white gap Reading</th>
<th>Hispanic-white gap Math</th>
<th>Mexican-white gap Reading</th>
<th>Mexican-white gap Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
</tr>
<tr>
<td>2002</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
</tr>
<tr>
<td>2004</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
</tr>
<tr>
<td>2006</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
<td>-0.02 (0.00)</td>
</tr>
</tbody>
</table>

Note: Estimated standard deviations are shown in parentheses. Estimated standard deviations are calculated using 2000 NAEP scores for (re)eeducational achievement levels for fourth-grade math, science, and English-language arts.
Steidl, and Freeman (2013), using data from the early 1990s through the late 2000s, found that gap trends were correlated with trends in racial school segregation, net of other time varying factors. Nonetheless, much of the research linking segregation to achievement gaps is either based on data from the 1960s and 1970s—when desegregation efforts led to large exogenous changes in segregation patterns—or is correlational in nature, which leaves some uncertainty as to whether segregation patterns are causally related to achievement gaps.

**Trends in Hispanic-White Achievement Gaps**

Despite the availability of NAEP-LTT data on a sizable Hispanic student population, there is relatively little research on patterns and trends of Hispanic-white achievement gaps. Reardon, Valentino, Kalogrides, Shores, and Greenberg (2013) show that there have been modest decreases in the Hispanic-white achievement gap during the last 20 years. Tables 28.1 and 28.2 present standardized achievement gaps calculated from NAEP data. The Hispanic-white reading and math gaps trends are similar to the black-white gap trends: the gaps decreased from the mid-to late-1970s to the mid-to late-1980s. Then increased through 1994 and have generally been declining since 1999. The Hispanic-white reading gap is generally similar in magnitude to the black-white reading gap. In math, however, the Hispanic-white gap has been consistently smaller than the black-white gap.

Data from Main NAEP also allow for the examination of achievement patterns of Mexican-origin students relative to non-Hispanic white students (Table 28.2). In general, the Mexican-white gaps are very similar to the overall Hispanic-white gaps (not surprisingly, since Mexican-origin students make up roughly two-thirds of Hispanic students).

**Trends in Socioeconomic Achievement Gaps**

Until recently, relatively little research has explicitly examined trends in what we might call the *socioeconomic status achievement gap*—the gap between children from lower-and higher-income or socioeconomic status families. Reardon (2011), however, estimated the income achievement gap (which he defined as the average test score gap between children from families at the 90th percentile of the family income distribution and children from families at the 10th percentile) from a set of nationally representative samples of students born from the early 1940s to 2001. He found that the income achievement gap has grown roughly 40 percent over the last few decades. These trends, he argues, likely result from the combination of rising income inequality and changes in patterns of families’ investments in their children since the 1970s, though more research is needed to fully understand these trends. More recent research suggests that the income achievement gap in school readiness may have started to decline in the last decade (Reardon & Portilla, 2014), though it is unclear if this represents a reversal of the trend Reardon (2011) described or simply a short-term anomaly.

In Table 28.1, we make an effort to examine socioeconomic achievement gaps using parental education level as a crude indicator of socioeconomic status. Specifically, we use NAEP-LTT data to compare the math and reading achievement of students whose parents have only a high school diploma to those with at least a (four-year) college degree.

The high school-college degree gaps in math were relatively stable between 1978 and 2008, ranging from roughly one-half to two-thirds of a standard deviation at age 13 and ranging from three-fifths to three-quarters of a standard deviation at age 17. In 2012, however, these gaps were higher than they have ever been in over 30 years. In reading, the parental education gaps have been generally smaller than those in math and show a little more of a trend over time. The reading gaps narrowed through the 1980s before widening again, so that the most recent estimates of the gap (roughly two-thirds of a standard deviation at ages 13 and 17) are now larger than they have ever been in the previous three decades.

**How Do Achievement Gaps Change as Children Progress Through School?**
Both racial and socioeconomic achievement gaps emerge prior to the start of kindergarten. For more information about the pre-kindergarten development of these gaps, see the chapter by Bassok & Loeb in this volume. Research findings on the development of achievement gaps as children progress through school come from two types of studies—those using longitudinal panel data on one or more student cohorts, and those relying on repeated cross-sectional data of the same cohort to infer developmental patterns (such as the NAEP studies). Under the respective assumptions that attrition (from longitudinal studies) is random and that changes in cohort composition are unrelated to achievement patterns (in repeated cross-sectional studies), both types of studies will provide unbiased estimates of the development of achievement gaps as students age. Regardless of whether longitudinal or repeated cross-sectional data are used to examine the development of gaps as students progress through school, the value of comparison of the magnitude of the gaps at different ages depends on the comparability of the test metrics used at each age. In fact, different test metrics lead to dramatically different conclusions regarding how achievement gaps change with age (compare, for example, Fryer & Levitt, 2004 and 2005, and Hanushek & Rivkin, 2006; see also Reardon, 2007). Gaps measured in pooled within-age standard deviation units are both more available and likely less sensitive to violations of the interval-scaled metric assumption than other approaches, so we focus here on studies that measure gaps in this type of metric.

**The Development of Black-White Achievement Gaps**

Almost all research on the topic concludes that the black-white achievement gaps grow during the school years, particularly in elementary school. The most commonly cited (and probably the best) contemporary evidence on the development of the black-white gaps in elementary school comes from ECLS-K, which includes kindergarten through eighth-grade assessment data on a nationally representative sample of students who were enrolled in kindergarten in the fall of 1998 (see Table 28.3). ECLS-K data show that the black-white gaps in both math and reading are sizeable at the start of kindergarten—about three-quarters and one-half of a standard deviation, respectively (Fryer & Levitt, 2004; Reardon, 2007; Reardon & Galindo, 2006). Measured in standard deviation units, these gaps widen slightly through kindergarten and first grade, and then widen more rapidly between first and Fifth grades, and then widen very slowly or not at all by eighth grade, by which time the math and reading gaps are about one full standard deviation (Benson & Borman, 2010; Reardon, 2007; Reardon & Galindo, 2006) (see Table 28.3).

Table 28.3 Estimated Achievement Gaps, Early Childhood Longitudinal Study: 1998 Kindergarten Cohort, by Group, Subject, and Grade
A number of other studies confirm that most of the growth in the black-white achievement gap occurs prior to fifth grade. Data from Study of Early Child Care and Youth Development (SECCYD) suggest the gap is large at the start of kindergarten, and grows in the early elementary grades (Murnane et al., 2006). NAEP-LTT data show that the black-white math gap (though not the reading gap) widens slightly from age nine to 13 (Ferguson, 1998; Neal, 2005; Phillips, Crouse, & Ralph, 1998). The development of NAEP reading and math gaps from age 13 to 17 is less clear—there is no consistent pattern in the size of the gaps. The gaps generally do not appear to widen much in this period (Ferguson, 1998), but these results are less certain because differential dropout patterns may bias the estimates of the gaps at age 17. NELS data likewise suggest that the gaps change relatively little following eighth grade (LoGerfo et al., 2006; Neal, 2006). Finally, state-level data from Texas and North Carolina are consistent with this finding; the gap in state test scores in those states grows relatively little in standard deviation units over the latter half of elementary school and middle school (Clotfelter et al., 2006; Hanushek & Rivkin, 2006).

The Development of Hispanic-White Achievement Gaps

The most detailed evidence on the development of Hispanic-white gaps comes from the ECLS-K, which includes a large sample (roughly 4,000 students) of Hispanic students that can be disaggregated by national origin, generational status, and English proficiency (Reardon & Galindo, 2006). In addition, because the ECLS-K study administered the math test orally in either English or Spanish, depending on students’ language proficiency, ECLS-K estimates of the Hispanic-white math gap are not biased by the changing English proficiency of the students.3

ECLS-K data (Table 28.3) indicate that the Hispanic-white math and reading gaps at the start of kindergarten are very similar in magnitude to the black-white gap; yet math and reading gaps decrease for Hispanic students during elementary school, while the black-white gaps widen during the same period (Fryer & Levitt, 2004 and 2005; Readon & Galindo, 2006). In math, the gap shrinks by 40 percent, from three-quarters to less than one-half of a standard deviation. Most of the narrowing of the gap occurs in kindergarten and first grade.
Evidence from other studies suggests the Hispanic-white achievement gaps continue to narrow, albeit slowly, through middle and high school. Data from North Carolina indicate that Hispanic students gain ground on white students in both math and reading from third through eighth grade, closing both gaps by over 0.10 standard deviations (Clotfelter et al., 2006).

Hispanic-white gaps do not develop similarly for all subgroups of the Hispanic student population. Using ECLS-K data, Reardon and Galindo (2006) provide a detailed description of the development of achievement gaps from kindergarten through fifth grade, by Hispanic subgroups (as defined by national origin and generational status). They find that Hispanic-white math and reading gaps appear to narrow during elementary school for all Hispanic subgroups, though the rate and pattern of this narrowing differs across subgroups. Reardon and Galindo speculate that because the rapid gains for Hispanics occur generally in the first two years of schooling, and because they are concentrated among recent immigrants and students with the lowest levels of English proficiency at the start of kindergarten, it is likely that much of the narrowing of the gap in kindergarten and first grade is attributable to the development of English language skills among these students. Because most other studies do not have sufficiently large samples of Hispanic subgroups (or do not collect data on national origin and generational status), we know relatively little about the development of achievement gaps by Hispanic subgroups in middle and high school.

The Development of Asian-White Achievement Gaps

There is relatively little detailed evidence regarding the development of Asian-white achievement gaps. Reardon and Galindo (2006) and Fryer and Levitt (2005) report the development of Asian-white gaps using ECLS-K data for the three-quarters of Asian-origin students proficient in oral English at the start of kindergarten.² These estimates likely overstate the average achievement of Asian-origin students, since the excluded students are generally from recently immigrated and lower-socioeconomic-status families. That said, the ECLS-K data (Table 28.3) show that Asian-origin students who are proficient in oral English at the start of kindergarten appear to have similar levels of math skills as white students through third grade, but better math skills in Fifth and eighth grades. In reading, the Asian-origin students in the sample have scores roughly two-to-three-tenths of a standard deviation higher than white students through first grade. This advantage narrows to nonsignificance by Fifth grade, but slightly increases to almost twenths of a standard deviation in eighth grade (Fryer & Levitt, 2005; Reardon & Galindo, 2006). Cheadle (2008) finds a similar reading advantage for Asian students through first grade, but the advantage narrows rapidly between second and third grade. He finds that the Asian-white gap does not grow during kindergarten, decreases during first grade, and widens in second and third grades. Table 28.3 includes our estimates of the Asian-white gaps computed from ECLS-K.

Given the heterogeneity of the Asian population—in terms of national origin, recency of immigration, socioeconomic status, and context of immigration—generalizations about Asian students as a monolithic group are as problematic as they are for Hispanic students. Certainly more detailed research on Asian-origin students’ achievement patterns would be useful.

The Development of Socioeconomic Achievement Gaps

Although considerable research documents the strong association between family socioeconomic characteristics and children’s cognitive development and school achievement, there is relatively little such research that provides a descriptive analysis of the development of the socioeconomic achievement gradient over time. Reardon (2011) finds that the income achievement gap changes little as children progress through school. One factor confounding such analyses is the fluid nature of socioeconomic status—unlike race or gender, socioeconomic characteristics of a family change over time, oft en quite dramatically.

In addition to the findings by Reardon (2011) on the patterns of cognitive development over time, we present here some new analyses using the ECLS-K and NAEP data and add gaps by parental education. Using an indicator of the highest level of education completed by either parent, we begin by examining trends in NAEP-LTT data. Averaging across all years for which we have NAEP data (see Table 28.1), we see that the average reading achievement gap between students from families with a high school diploma and those with a four-
year college degree widens from 0.48 standard deviations at age 13 to 0.55 standard deviations at age 17. The average math gap widens from 0.59 standard deviations to 0.67 standard deviations from age 13 to 17. Of course, if low-achieving students from less educated families leave school before age 17 at higher rates than similarly low-achieving students from families with college-educated parents, then these numbers may underestimate the rate of increase in the socioeconomic gaps as children progress through school.

Data from longitudinal studies such as ECLS-K and NELS also show that socioeconomic gaps widen somewhat as students progress through school, except during the first few years of schooling, when they appear to narrow modestly (Reardon, 2011). Our analysis of ECLS-K data suggests that both math and reading achievement gaps narrow by roughly 10 percent during the first two years of schooling, but then widen slowly through eighth grade (Table 28.3).

Do Achievement Gaps Grow Differently Across the Achievement Range?

Most studies examining achievement disparities between groups focus on differences in mean achievement. There are, however, important reasons to examine the disparities across the full distribution of test scores. For example, underlying the debate regarding affirmative action in admissions to highly competitive colleges is the fact that black and Hispanic students are dramatically underrepresented in the upper end of the achievement distribution. Neal (2006, see Figures 2a–2d) shows that roughly 5 percent of black students aged 13–17 years old in the 1990s had math scores in the top quartile of the white math score distribution. This means that black students are underrepresented by 80 percent in the top quartile of the distribution, a finding that has enormous implications for black students’ access to elite colleges and employment in jobs with the highest skill demands (and the highest pay). Such patterns suggest the importance of investigating not only differences in the black and white test score distributions, but also of investigating when and how such differences emerge.

Answering this question turns out to be more complex than it would seem, however, because any comparison of the magnitude of gaps or differences in growth rates relies on the assumption that the test metric used is interval-scaled. Clotfelter, Ladd, and Vigdor (2006) investigate whether the gap in scores between the 90th percentiles of the black and white test score distributions grows or narrows faster than the gap between the 10th percentiles of the distributions. They find that in math, racial test score gaps measured in standard deviation units generally narrow from third to eighth grades at the 10th percentiles of the score distributions, and widen at the same time at the 90th percentiles of the distributions. They find no such pattern for reading. Clotfelter, Ladd, and Vigdor interpret the math pattern as potentially a result of accountability pressures, arguing that the compression of the gap at the low end of the test score distribution is a result of policies that push schools to reduce the percentage of students scoring below certain thresholds. Likewise, they view the expansion of the gap at the high end as a result of the diversion of resources away from high-achieving minority students (because such students are in schools with many low-achieving students). While this is a plausible explanation, it is also possible that the results are an artifact of the tests used to measure the gaps. If the third- and eighth-grade tests are not both scored in interval-scaled metrics, and if the eighth-grade test metric is more sensitive to variation at the high end of the distribution than is the third-grade test, then the pattern they find would be observed in the absence of any true difference in the rate of the gap growth.

Likewise, determining whether achievement gaps grow faster or slower between initially high-and low-achieving students relies on the assumption that test scores are interval-scaled. In addition, measurement error in test scores will also tend to bias such estimates, because conditioning growth rates on scores measured with error will systematically bias estimates of differences in growth rates (Hanushek & Rivkin, 2006; Reardon, 2007). Relatively little empirical research has attempted to systematically address the question of whether achievement gaps within a cohort grow or narrow differentially across the range of skill distribution. What research there is has focused exclusively on the black-white gap and generally has not adequately addressed the complexities described here. As a result, there is little we can say with certainty to this point.

How Much of Racial/Ethnic Achievement Gaps Can be Explained by
Socioeconomic Status?

A relatively common question addressed in studies of racial/ethnic achievement gaps (particularly the black-white gap) is the extent to which the observed gaps can be explained by socioeconomic differences between the groups. Using ECLS-K data, Fryer and Levitt (2005) show that socioeconomic factors explain almost all (85 percent) of the black-white math gap and all of the reading gap at the start of kindergarten (in fact, they find that the black students score higher in reading than white students of the same socioeconomic status). By third grade, however, they find that the same socioeconomic factors account for only about 60 percent of both the math and reading black-white gaps. This finding suggests that socioeconomic factors explain, in large part, the black-white differences in cognitive skills at the start of formal schooling, but do not account for the growth of the black-white gap as children progress through elementary school. This observation has significant implications for understanding the role of schooling in producing or exacerbating achievement gaps.

The Fryer and Levitt analysis is notable for being the only one that shows that the black-white gap at kindergarten entry can be almost completely accounted for by socioeconomic differences between black and white students. Earlier studies typically found that socioeconomic factors explain roughly half of the black-white gap at kindergarten entry, though these results were based on samples of children who are disproportionately poor (Brooks-Gunn, Klebanov, & Duncan, 1996; Phillips, Brooks-Gunn et al., 1998). Other recent studies, however, show that socioeconomic factors typically account for even less—between 25 and 40 percent—of the black-white math and reading gaps through middle and high school. Using the same ECLS-K data as Fryer and Levitt but a different version of the test score metric (the unstandardized scale scores), Murnane and colleagues (2006) report that the socioeconomic differences between white and black students accounts for only one-third of the math gap and 15 percent of the reading gap by third grade (though they find that socioeconomic factors fully account for the black-white gap in kindergarten). Similarly, using the SECCYD data, they find that socioeconomic status accounts for one-third of the math and one-quarter of the reading black-white gaps in third grade (Murnane et al., 2006). Phillips, Crouse, and Ralph’s (1998) analysis of Prospects and NELS data shows SES explains about 35 to 40 percent of black-white achievement gaps in math, reading, and vocabulary across the span of second through 12th grade. North Carolina data show similar patterns, with socioeconomic factors (as well as region of the state and urbanicity) accounting for about 35 percent of the math and reading gaps from third through eighth grades (Clotfelter et al., 2006).

Yet, recent research suggests that socioeconomic factors may explain an even larger share of the black-white achievement gap than estimated by Fryer and Levitt. Rothstein and Wozny (2013) argue that parent-reported current annual income (which is what Fryer and Levitt and most studies use) is a noisy measure of families’ long-run income; the use of a noisy measure of long-run income leads to underestimation of the extent to which race/ethnicity achievement gaps are explained by socioeconomic status. When Rothstein and Wozny take long-run income into account, they estimate that the black-white math gap in third grade is 0.15 standard deviations, considerably smaller than Fryer and Levitt’s (2006) estimate of 0.38 standard deviations.

Other research using data from the National Longitudinal Survey of Youth suggests that the black-white achievement gap in adolescence can be entirely explained by models that account for multigenerational measures of socioeconomic status (Mandara, Varner, Greene, & Richman, 2009). Mandara and colleagues found that, conditional on parental and grandparental socioeconomic status and parenting practices, black students actually have slightly higher test scores than white students. Thus, research accounting for long-run parent income (e.g., Rothstein & Wozny, 2013) or multigenerational socioeconomic status (e.g., Mandara et al., 2009) tends to explain a much greater proportion of black-white achievement gaps than does research using current, single-generational measures.

The evidence regarding the extent to which socioeconomic factors account for Hispanic-white achievement gaps is less mixed: research consistently finds that these gaps are largely or entirely explained by socioeconomic status. For example, using ECLSK data, Fryer and Levitt show that socioeconomic factors account for 75 to 85 percent of the Hispanic-white gaps in kindergarten and 85 to 100 percent of the gaps in third grade (Fryer & Levitt, 2006). One explanation for this different pattern may be that the Hispanic-white gaps in kindergarten are partly due to Hispanic students’ lower levels of English proficiency; as students progress through school, their English skills improve, and the gaps narrow (see Table 28.3), so that socioeconomic factors explain an increasing proportion of the gaps. Similarly, Reardon and Galiano (2006,
tables B11, B12) show that Hispanic-white gaps, conditional on socioeconomic status, narrow from kindergarten through fifth grade, while the corresponding black-white gaps widen at the same time. North Carolina data show a similar pattern: Hispanic-white socioeconomic-adjusted math and reading gaps are small (one-tenth of a standard deviation) in third grade, nonexistent by fifth grade, and reverse sign by eighth grade, when Hispanic students score higher, on average, in both math and reading than socioeconomically similar white students (Clotfelter et al., 2006).

Do Achievement Gaps Grow Between or Within Schools?

A central question in understanding test score gaps is the extent to which such gaps can be attributed to differences in average school quality between schools attended by students of different racial, ethnic, or socioeconomic groups. If, for example, black students attend, on average, lower quality schools than do white students, we would expect the between-school component of the black-white achievement gap to grow over time. If black and white students receive unequal instructional opportunities when attending the same schools, we would expect the within-school component of the black-white gap to grow over time. Of course, it is difficult to disentangle the effects of school quality from the sorting processes that produce racially and socioeconomically segregated schools and that may result in lower-ability students, regardless of race or socioeconomic status, attending schools that have lower proportions of white or middle-class students. Likewise, it is not clear that differences in achievement gains can be attributed solely to schooling processes (particularly given the evidence that the gaps predate kindergarten; see Fryer & Levitt, 2004; Lee & Burkham, 2002; and Reardon & Galindo, 2006), given unequal family resources, neighborhood context, and opportunity structures (which may lead to unequal motivation even in the presence of equal home and school resources). Nonetheless, an understanding of the relative contribution of between-and within-school factors, as well as of family background and out-of-school social context, is essential for determining the appropriate policy remedies for the gaps.

Despite the importance of disentangling between-and within-school patterns in the growth of achievement gaps, there is relatively little clear evidence on this point. Fryer and Levitt (2005) find that the black-white gap is small in kindergarten, net of family socioeconomic characteristics, but grows from kindergarten through third grade (using ECLS-K data)—a pattern that suggests that observable family background characteristics are not solely responsible for the post-kindergarten growth in the gap. However, they also find that the black-white gap grows through third grade even between black and white students attending the same school, a finding they interpret to mean that between-school differences in school quality do not account for much of the growth of the black-white gap during elementary school (but see Hanushek & Rivkin, 2006; see also Reardon, 2007, for an explanation of discrepancies between Fryer & Levitt’s and Hanushek & Rivkin’s analyses).

Another approach to studying whether gaps develop within or between schools is to investigate the extent to which beneficial instructional practices and teacher attributes are differentially distributed across different groups of students within and between schools. For example, using national data (ECLS-K), Desimone and Long (2010) found that less experienced teachers were more likely to be assigned to minority students and students from low-income families. These patterns are certainly partly due to differences among schools in average teacher experience levels, differences that are associated with school racial composition. However, more recent research provides evidence that teacher sorting occurs within schools as well as between schools. Kalogrides and colleagues found that novice teachers are assigned to lower-achieving, minority, and poor students within the same school, potentially exacerbating achievement gaps (Kalogrides & Loeb, forthcoming; Kalogrides, Loeb, & Béteille, 2013). Another possible source of widening achievement gaps is growing racial segregation (Berends & Perialoza, 2010; Condron, Tope, Steidl, & Freeman, 2013). In sum, the literature suggests that a number of factors—both within and across schools—likely contribute to racial achievement gaps.

What Are the Labor Market Effects of Achievement Gaps?

430
From a labor market perspective, achievement disparities are important primarily because test score disparities in elementary and secondary school are highly predictive of corresponding disparities in subsequent labor market outcomes. Data from the Current Population Survey (CPS) from 2011 and 2012 show that the median black and Hispanic male full-time workers earn 76 percent and 67 percent of what the median white full-time male worker earns. For females, the median black and Hispanic full-time workers earn 84 percent and 73 percent of their white counterparts. These wage differentials have remained stable for the past decade.

A sizeable body of research has investigated the extent to which these wage disparities are attributable to differences in cognitive skill obtained prior to entering the labor force (i.e., in childhood, elementary school, and secondary school), typically measured by cognitive achievement test scores. In general, this research finds that roughly one-half to all of the male black-white wage gap can be accounted for by black-white differences in human capital, as proxied by scores on the Armed Forces Qualification Test (AFQT) when individuals were near completing high school (Bollinger, 2003; Carneiro, Heckman, & Masterov, 2003; Neal & Johnson, 1996). Although there is much less evidence regarding the extent to which test score differences account for Hispanic-white wage gaps, evidence from NLSY suggests that test score differences account for virtually all of the male Hispanic-white wage gap (Carneiro et al., 2003). These findings suggest that factors such as racial discrimination in the labor market are not responsible for much, if any, of the black-white wage differential, a conclusion supported by other research that compares earnings differentials among different cohorts. For example, O’Neill (1990) reviews possible explanations for the decrease in the black-white wage differential during the 20th century, and finds that differences in the educational and workplace experiences of black and white men are much more important in explaining the wage gap between the races than is labor market discrimination. Examining hourly wages six years after high school graduation, Murnane, Willett, and Levy (1995) find that cognitive skills predict labor market outcomes more in the mid-1980s than they did in the late-1970s, suggesting the increasing importance of cognitive skills in wage determination.

It is worth noting, however, that not all gains in reducing black-white wage inequality during the 20th century can be attributed to narrowing skill gaps (O’Neill, 1990). For instance, the relative earnings of blacks increased during the 1960s and 1970s—arguably due to governmental intervention in reducing employment discrimination—for cohorts who had already completed schooling (Donohue & Heckman, 1991). Likewise, some research suggests that labor market discrimination remains an important factor in labor market outcomes (see, e.g., Grodsky & Pager, 2001, and Pager, 2003), though even this research acknowledges that the majority of the black-white wage differential can be attributed to black-white differences in human capital obtained prior to entering the labor market.

The finding that human capital differences account for a large portion of black-white wage gaps suggests that efforts to reduce black-white achievement disparities by the end of adolescence may substantially reduce subsequent black-white wage gaps, though such a reduction would require decades to take full effect. Even an immediate elimination of the gaps in human capital among adolescents would not, presumably, reduce wage disparities among cohorts already in the labor force. As a result, wage differentials among older cohorts may persist for decades, until these cohorts age out of the labor force.

In addition to concerns regarding the magnitude of the differences in mean test scores among individuals of different racial groups, a number of researchers have called attention to the effects of racial disparities at the upper end of the achievement distribution (Hanushek & Rivkin, 2007; Neal, 2006). Moreover, some evidence indicates that the increase in the returns to education in the 1980s was largest for those in the top quartile of the achievement distribution (Heckman & Vytlacil, 2001). Because whites are substantially overrepresented in the highest quartile of the achievement distribution, this pattern suggests that racial disparities at the top of the achievement distribution have become increasingly salient in shaping labor market inequality.

Finally, although much of the economics literature focuses on the role of cognitive skills in explaining wage gaps, there is an emerging emphasis on the role of “noncognitive” skills (e.g., motivation, self-control, perseverance) on wages. For example, a recent study using data from the Perry Preschool experiment examined the extent to which both cognitive and noncognitive benefits from the experimental condition (focused on children from low-income families) affected income at age 27 (Heckman, Pinto, & Savelyev, 2012). The analyses revealed that, for males, both cognitive and noncognitive factors significantly predicted differences between the treatment and control groups in monthly income at age 27 and whether the individual was employed at age 40. In addition, treatment effects in noncognitive factors associated with reduced externalizing...
problem behaviors at ages seven to nine (e.g., lying, disrupting class, acting aggressive to peers) were associated with treatment differences in other adult outcomes such as drug use and arrests, factors that are likely undesirable to employers. This evidence suggests that noncognitive skills have an effect on later income, independent of cognitive skills. Moreover, other evidence suggests that noncognitive skills are associated with achievement (Duckworth & Seligman, 2005) and socioeconomic status (Moffitt et al., 2011; Noble, McCandliss, & Farah, 2007), suggesting that noncognitive skill gaps in childhood may contribute to academic achievement gaps as well. Thus, interventions might consider a focus on reducing noncognitive gaps to partially address both cognitive gaps and later wage gaps (Lindqvist & Vestman, 2011).

Conclusion

In this review, we have attempted to summarize the state of knowledge regarding racial/ethnic achievement gaps and to suggest areas in which more research is needed. Despite the complexity of answering questions about achievement gaps due to the need to rely on imperfect measures of cognitive skills, several key patterns are evident from our summary of the research.

First, racial/ethnic achievement gaps are narrower now than they were 30–40 years ago; this is particularly true for the black-white achievement gaps. Second, these gaps remain quite large today, ranging from 0.5 to 1.0 standard deviations. Third, the black-white gap appears to widen during the school years—particularly in early elementary school—in ways that are not explained by socioeconomic family background characteristics, a pattern that suggests that schooling appears to contribute to the growth of the gaps. The same patterns are not found for Hispanic-white and Asian-white disparities, however; for these groups, socioeconomic differences account for a large portion of the gaps, and processes of second-language acquisition appear to contribute to a narrowing of the gap as children progress through school. Fourth, achievement disparities have large effects in the labor market and explain a large portion of racial/ethnic income disparities.

Less clear in extant research are the processes and mechanisms that produce racial and ethnic achievement disparities. Although it is clear that family background and schooling each play some role in the development of achievement gaps, we do not have good evidence on exactly how—or how much—each contributes. Nor do we have any good evidence regarding the extent to which these processes may vary among racial or ethnic groups. Most importantly, without a good understanding of the mechanisms that produce these gaps, we have very little evidence regarding how we might reduce them. In addition to documenting the magnitude, trends, and development of achievement gaps, we need more and better research regarding the effectiveness of social and educational policy to reduce them.

Notes

1. For more information on NAEP, see http://nces.ed.gov/nationsreportcard/about/.
2. Main NAEP also includes assessments in writing, science, and other subjects. We focus in this chapter on the math and reading NAEP assessments.
3. We report the NAEP estimates of achievement gaps in standard deviation units, using the population standard deviation for each specific grade-subject-age/year combination. There are, of course, a number of different metrics in which achievement gaps can be measured, including differences in average scale scores, differences in pooled standard deviation units, and differences in average ranks, the latter of which are the basis for a set of "metric-free" gap measures (Ho & Haertel, 2006). The empirical literature on achievement gaps has generally relied primarily on standardized gap measures (see, for example, Clotfelter, Ladd, & Vigdor, 2006; Fryer & Levitt, 2004 and 2005; Grissmer, Flanagan, & Williamson, 1998; Hedges & Nowell, 1999; Neal, 2005; Phillips, Brooks-Gunn, Duncan, Klebanov, & Crane, 1998; and Reardon & Galindo, 2006), while the "metric-free" measures have been less widely used (for recent examples of their use, see Ho & Haertel, 2006; Neal, 2005; Reardon, 2007; and Robinson & Lubinski, 2011). In addition, some of the literature on achievement gaps has relied on (un-standardized) mean differences in test scores (Hanushek & Rivkin, 2006; LoGerfo, Nichols, & Reardon, 2006; Murnane, Willett, Buh, & McCartney, 2006). Although inferences about achievement gaps can be dependent on the choice of a test metric (see, for example, Bond & Lang, 2013; Reardon,
In addition, we might expect that—to the extent that any of the current wage differentials is due to employers’ “statistical discrimination” practices (i.e., employers’ use of race or education levels as a proxy for unobserved skill levels believed to differ systematically between racial groups)—the experiences of the current U.S. Hispanic population. AFQT scores (Bollinger, 2003; Carneiro et al., 2003).

When comparing Hispanic-white achievement gaps across cohorts or grades, some caution is required. Given the dramatic increase of the Hispanic student population over the past few decades, the composition of Hispanic students tested in different grades and different years may vary greatly, and so any comparison across cohorts or grades may be confounded with the changing composition of the Hispanic student population.

Examples of such studies include those using panel data from nationally representative samples—such as the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K) (see http://www.nces.ed.gov/ecls), the National Education Longitudinal Study (NELS) (see http://www.nces.ed.gov/surveys/nels88), Prospects: TheCongressionally Mandated Study of Educational Growth and Opportunity, and High School and Beyond (HSB) (see http://www.nces.ed.gov/surveys/hsb)—and those drawn from state administrative data sources in states like North Carolina, Texas, or Florida, each of which has administrative data systems allowing tracking of individual student test scores over multiple years (Cloftelter et al., 2006; Hanushek & Rivkin, 2006).

The assumption of no cohort change may be particularly problematic, however. For example, a sample of nine-year-old students drawn in 1991 and a sample of 17-year-old students drawn in 1999 may not represent exactly the same cohort population, since the 1999 sample would include some students not in the 1991 population (e.g., those born in 1982 who immigrated to the United States between 1991 and 1999), and the 1991 sample would include some students not in the 1999 population (e.g., those who dropped out of school between age nine and age 17).

Because the ECLS-K reading test was administered only in English, students not proficient in oral English were not assessed in reading; as a result, estimates of the Hispanic-white reading gap are typically generalizable only to the subsample of Hispanics proficient in oral English at the start of kindergarten (see, for example, Fryer & Levitt, 2004 and 2005; and Reardon & Galindo, 2006). In math, Hispanic students not proficient in English took the test in Spanish. Although there is evidence that the scores of some Hispanic students are biased downward because they took the math test in English rather than Spanish, this bias likely affects only Hispanic students whose English proficiency was just above the threshold level required for them to take the test in English; it is unlikely to substantially affect the overall pattern of Hispanic math scores (Robinson, 2010).

The ECLS-K math test was administered only in English or Spanish (and the reading test was administered only in English), so the roughly 22 percent of Asian-origin students in the sample who were not proficient in oral English did not take the tests in the fall of kindergarten.

The comparison across cohorts relies much less on the assumption of interval scaling, since it is possible to compare the full test score distributions across cohorts. See, for example, Hedges and Nowell (1999) and Ferguson (1998).

One possible reason why Fryer and Levitt’s model explains more of the black-white gap in elementary school than do other studies may be that they use a somewhat larger list of covariates than do other studies, including age, birth weight, gender, number of children’s books (and its quadratic term), mother’s age at first birth, WIC receipt, and a composite socioeconomic status variable that includes family income, parental education, and parental occupation measures (Fryer & Levitt, 2006).

Murnane et al.’s results differ substantially from those of Fryer and Levitt likely because Murnane et al. report the predicted black-white gap from a random-coefficient growth model while Fryer and Levitt report results from a series of repeated cross-sectional covariate adjustment models (Fryer & Levitt, 2006; Murnane et al., 2006). We prefer the Fryer and Levitt results because they rely less on assumptions regarding interval scaling of the test metric and the stable effects of socioeconomic factors over time.

With regard to wage gaps for women, the evidence is less clear because of differential selection into the labor force among women. Among women in the labor force, however, black and Hispanic women earn, on average, the same or more than white women after adjusting for AFQT scores (Bollinger, 2003; Carneiro et al., 2003).

Note that the NLSY Hispanic sample is made up of Hispanics living in the United States in 1979, so this finding may not be representative of the experiences of the current U.S. Hispanic population.

In addition, we might expect that—to the extent that any of the current wage differentials is due to employers’ “statistical discrimination” practices (i.e., employers’ use of race or education levels as a proxy for unobserved skill levels believed to differ systematically between racial
groups)—an immediate reduction in achievement gaps may not yield immediate wage gap reductions even among the youngest cohorts, because it may take time for employers to adjust their prior beliefs about racial differences in human capital. However, Altonji and Pierret (2001) find little evidence for statistical discrimination on the basis of race, net of educational attainment. Moreover, they find also that although employers may statistically discriminate on the basis of employees' educational attainment in assigning initial wage levels, employees’ productivity plays a much larger role in wage determination as the employee accrues more experience—employers appear to adjust their prior beliefs, at least with regard to their own employees, as they gain more information. These findings suggest that the impact of a reduction in the black-white skill gap may have a relatively immediate effect on wage gaps among cohorts for whom the skill gap is reduced.

References


Early Childhood and the Achievement Gap

DAPHNA BASSOK AND SUSANNA LOEB

By the time children enter kindergarten socioeconomic and racial school-readiness gaps are deeply entrenched. Data from the 2010 Early Childhood Longitudinal Study (ECLS-K), a large, nationally representative survey, show that at kindergarten entry, the average reading and math scores of children from high socioeconomic backgrounds are a full standard deviation higher than those of low-income children (Reardon, 2014). Significant differences in cognitive assessment scores are also evident between racial groups, with white students scoring about half of a standard deviation higher than black children on a math assessment and about a third of a standard deviation higher on a test of reading (Reardon, 2014). Study after study highlights early childhood gaps across both cognitive and noncognitive measures (Duncan & Magnuson, 2011; Isaacs, 2012) which surface as early as nine months and widen throughout early childhood (Fernald, March-man, & Weisleder, 2013; Halle et al., 2009; Shonkoff & Phillips, 2000).

Three critical and interrelated reasons support focusing policy and research attention on the years prior to children’s school entry. First, environmental factors have tremendous impact on developmental trajectories. While environments affect development throughout the life course, children’s experiences in the early years have a disproportionately large impact relative to experiences during school-age years and later (Heckman, 2006). Second, children’s skills and knowledge when they are young are strongly related to their later life outcomes (Chetty et al., 2011; Duncan et al., 2007). That is, children who lag behind early are likely to continue doing so throughout their schooling and beyond. Finally, an increasing body of research has demonstrated that high-quality interventions targeted either at young children or their parents can make meaningful long-term impacts (Blau & Currie, 2006).

The growing consensus around these three points is reflected in the substantial increases in U.S. public investment in early childhood over the past decades (Reynolds, Rolnick, Englund, & Temple, 2010). Increases in state spending for preschool have been particularly pronounced, rising from $2.4 to $5.1 billion between 2002 and 2012 (Barnett, Carolan, Fitzgerald, & Squires, 2012). Forty states now offer state-funded preschool programs, and eight serve over 50 percent of their four-year-old children. The federal government has also taken on a heightened role in early childhood education. Through its Race to the Top—Early Learning Challenge, the federal government has competitively allocated roughly $1 billion to 20 states since 2011, with the goal of improving access to high-quality early learning experiences.

The purpose of this chapter is to synthesize the ballooning body of research evidence on the impacts of early childhood interventions, with an explicit focus on their potential for ameliorating achievement gaps.

The Unique Importance of Early Childhood Environment

Recent advancements in the neurobiological and social sciences have vastly expanded our understanding of development in the early years of life (Collins, Maccoby, Steinberg, Mavis, & Bornstein, 2000; Knudsen, Heckman, Cameron, & Shonkoff, 2006; Nisbett et al., 2012; Shonkoff, Richter, Gaag, & Bhutta, 2012). While genetic make-up explains some developmental trajectories, environmental factors—including economic well-being, physical surroundings, communication, and nurturing—interact with children’s genetic endowment, affecting their cognitive, social, and emotional development.

Hart and Risley’s (1995) seminal study demonstrates the strong association between children’s early environments and their development. The authors document that young children of welfare recipients heard
approximately 616 words per hour compared to roughly 2,153 words per hour in the homes of higher-income children. Extrapolating from hourly discrepancy suggests that by age four, children in high-income homes have likely heard roughly 30 million more words than have their low-income peers. Further, the amount and type of language exposure children experience closely tracks their language development early in life. The study does not establish a causal link between language exposure and children’s development; it does not definitively rule out other socioeconomic factors or even genetic explanations for differences across groups. However, it shows that environmental factors differ widely across groups and may substantially influence development.

Indeed, an expansive literature demonstrates associations between a whole host of environmental factors and children’s early development. While a detailed review of this body of work is beyond the scope of this chapter, we note that researchers have found positive relationships between child outcomes and breast-feeding (Kramer, Aboud, Mironova et al., 2008; Kramer, 2010), exposure to enriched language environments (Fernald et al., 2013; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991), and various home literacy practices (Rodriguez et al., 2009; Sénéchal & LeFevre, 2002). Conversely, divorce and single parenting (Amato, 2010; McLanahan & Percheski, 2008), maternal depression (Knitzer, Theberge, & Johnson, 2008) and early exposure to stressful environments (Shonkoff, Garner et al., 2012) are among factors that are negatively correlated with child outcomes.

Given the interrelatedness of these factors, and the unclear directionality among them, the relationships presented so far should be interpreted as associations and not necessarily causal. That said, increasingly rigorous social science research which leverages experimental and quasi-experimental approaches demonstrates that early childhood poverty (and the many embedded processes) meaningfully impact life trajectories (Yoshikawa, Lawrence, & Beardslee, 2012). Further, studies leveraging longitudinal data show that early childhood poverty in particular, is a strong predictor of various later life outcomes—much more so than resources in middle childhood or later in life (Duncan & Brooks-Gunn, 1997; Duncan, Yeung, Brooks-Gunn, & Smith, 1998; Duncan, Ziol-Guest, & Kalil, 2010).

The Link Between Early Abilities and Later Achievement

Young children’s cognitive and social abilities are strongly related to their later school and life outcomes (Chen, Lee, & Stevenson, 1996; Claessens, Duncan, & Engel, 2009; Claessens & Engel, 2013; Luster & McAdoo, 1996). For instance, Duncan et al. (2007) leverage six large longitudinal datasets to show that children’s cognitive skills at school entry, particularly their early math skills, are strong predictors of their later academic performance. Relatedly, in their meta-analysis of eight national studies of the black-white achievement gap, Phillips, Crouse, and Ralph (1998) found that the black-white test score gap at the end of high school could be reduced by at least one-half if the gap were eliminated at school entry.

Other studies have measured the relationship between early childhood assessment scores and adult outcomes. For instance, Chetty et al. (2011) demonstrate that kindergarten test scores predict earnings between ages 25 to 27, college attendance and college quality, and an index of other adult outcomes such as home ownership, marriage, and savings. These results align with older results from Baydar, Brooks-Gunn, and Furstenberg (1993) who followed over 200 children born to black, teenage mothers in Baltimore from birth to approximately age 20 and found that early childhood assessments strongly predict illiteracy 15 years later.

The Benefits of Intervening Early

One of the most celebrated early childhood interventions is the High/Scope Perry Pre-school experiment, which randomly assigned 123 at-risk, low-income black children to either a high-quality, two-year preschool intervention or a control group. Researchers have followed these children from age three through age 40 and have reported dramatic and long-lasting effects (Schweinhart et al., 2005). For instance, while 65 percent of program participants graduated from high school, only 45 percent of the control group did; for females the proportions were 84 and 32 percent, respectively. At age 40 the median annual earnings of program participants were $20,800, compared with $15,300 for the control group. Participants were also more likely be
employed, raise their own children, and own a home or a car, and they were far less likely to experience arrests or utilize drugs. Researchers estimated that the Perry Preschool program had a return of approximately $258,888 per child, or $17.07 per dollar invested. About three-quarters of this return went to the public through savings on crime, education, and welfare, as well as increased taxes. While involvement in a highly publicized experiment may have influenced the choices and outcomes of program participants, it is unlikely that these large, persistent effects are being driven only by heightened attention. Intensive interventions early in life appear to meaningfully impact the later success of poor children.

A full evaluation of the potential of early investment in child development requires a comparison of the returns of these types of interventions to those achieved through other avenues. Carneiro and Heckman (2003) compare the impact on life outcomes of early childhood interventions to those of school improvement efforts (e.g., class size reduction and teacher salary increases), adolescent interventions (e.g., mentorship and dropout prevention programs), job training programs, and college tuition breaks. They conclude that the most powerful drivers of adult outcomes are determined by environmental factors during early childhood. Heckman makes the case that "skills beget skills" and that later remediation of deficits is far more expensive than intervention during the first years of life (Heckman, 2008). Heckman's work is informed by a growing body of social science intervention research as well as by a greater understanding of brain development over the life course. In particular, we now understand that early childhood (between ages zero and five), is a period of extremely rapid brain development in which genes and environmental interactions are uniquely relevant for shaping life trajectories (Knudsen et al., 2006; Shonkoff, 2010).

Understanding Achievement Gaps

The findings just discussed provide evidence that interventions aimed at changing children’s environments—particularly interventions targeted towards low-income and minority children—are likely to meaningfully affect achievement gaps, both school entry and throughout life. Before turning to a summary of intervention literature, we examine another heavily researched possibility, namely that both socioeconomic and racial gaps are driven by genetic rather than environmental differences. Isolating the role of the environment is difficult because most children are brought up by their biological parents, who both determine their children’s genetic characteristics and help create their environment. In this section we briefly summarize the literature on the role of genetic and environmental factors in influencing child development, and stress recent discoveries that suggest important interactions between nature and nurture.

Scarr and Weinberg (1978) one of the seminal works in this field, measured the extent to which typical estimates of the importance of environmental factors are biased by the omission of genetic factors. They compared a sample of biological families to families of children adopted in the first few months of life, reasoning that because adopted children do not share their adoptive parents’ genetic characteristics, any correlations between environmental characteristics and child outcomes represent unbiased environmental effects. They concluded that for their sample of adolescents, ages 16–22, from working to upper middle-class families, environmental factors were negligible in explaining differences in IQ scores. These results have been replicated repeatedly across various samples and outcome measures, including cognitive and behavioral outcomes, as well as educational attainment (Plomin & Petrill, 1997; Plug & Vijverberg, 2003).

Many researchers, however, have taken serious issue with the notion that child development is so deterministically driven by genetic factors (Baumrind, 1993; Collins et al., 2000; McClearn, 2004; Nisbett et al., 2012; Sameroff, 2010). The key criticism is that the studies on which these conclusions are based assume that nature and nurture affect childhood development independently and aim to isolate each ones’ contribution ignoring important interaction effects. Duyme, Dumaret, and Tomkiewicz (1999) considered a sample of low-IQ children who were abused and neglected as infants and were adopted between ages four and six. Adoptive parents were defined as low-, medium-, or high-SES based on the father’s occupation. When the children were 13.5 on average, their IQs were reassessed. All children showed growth, but the magnitude of IQ change varied with parents’ SES, with the low-SES group raising their IQ by 7.7 and the high-SES group experiencing an increase of 19.5 points. These differing effects imply that highly enriched environments are likely to have significant effects on child outcomes.
In recent years the field of epigenetics, the study of how environmental experiences interact with genetic composition, has grown rapidly, demonstrating clear interdependencies (Meaney, 2010). A number of recent quasi-experimental studies also provide strong support for the potential of environmental interventions by leveraging large exogenous shifts to children’s environments (Gould, Lavy, & Paserman, 2004; Santavirta, 2012). For example Gould et al. (2011) exploit exogenous variation in the living conditions of children airlifted into Israel in 1949 and show that children who, by chance, were placed in better initial conditions had improved adult outcomes including educational attainment and employment, as did their children.

Explaining Group Differences

Until now, we have concentrated on the role of genetic and environmental factors in explaining differences among individuals instead of differences between racial or ethnic groups and across income groups. If group gaps are driven by environmental differences, then policymakers interested in narrowing the gap might pursue approaches focused on economic factors. If, on the other hand, there is something specific to particular groups that generates disparities, other policy approaches might be warranted.

One approach to parsing out these two issues for racial groups is to measure how much of the achievement gap remains after controlling for environmental factors, such as poverty levels and family structure (Duncan & Magnuson, 2005). Using data from the Early Childhood Longitudinal Study, Fryer and Levitt (2004) report that controlling for a small set of observable background characteristics the entire black-white test score gap essentially disappears for entering kindergarteners. The same set of controls eliminates the majority of the Hispanic-white gap as well. Yet, even though much of this initial gap can be explained by a small number of covariates, by the time children are in third grade, the black-white achievement gap has grown, and the same set of controls (as well as a number of other covariates) do not explain away the gap (Fryer & Levitt, 2006). Further, studies vary in the extent to which they can explain away early achievement gaps through the inclusion of environmental controls. Efforts to replicate kindergarten-entry results in other large datasets show that controlling for a very similar set of socioeconomic controls did not eliminate the black-white test score gap in either math or early literacy skills (Burchinal et al., 2011; Murnane et al., 2006).

Critics of the environmental argument argue that excluding a true measure of genetic background from models of the achievement gap would lead to an omitted variable bias because genetic characteristics could be related to both environment and children’s achievement. Despite these concerns, reviews of the literature have concluded that there is very little evidence to support the notion that significant portions of the achievement gap between groups (i.e., the black-white gap) can be explained by genetic factors (Dickens, 2005; Nisbett et al., 2012). Many of the most compelling arguments against the genetics hypothesis are found in studies that demonstrate powerful effects of environmental factors, on child outcomes (e.g. Schweinhart et al., 2005). Moreover, Fryer and Levitt (2004) find that controlling only for children’s age and gender, the gap between black and white infants on a developmental assessment is only 0.06 of a standard deviation, and is even smaller for Hispanics. This gap is quite small compared to the one observed among older children in other studies, and is therefore damaging to the genetics argument. The authors conclude that either genetic factors are not the primary drivers of racial achievement gaps or that genetic influences emerge systematically after age one.

Certainly, we have more to learn about the interplay among genetics, environmental factors, and child development. Even if genetic differences among racial or ethnic groups were to play a role in childhood development, these differences need not imply that environmental interventions would be ineffectual. Given the strong success of certain early childhood interventions in improving the well-being of poor, black and Hispanic children, altering environmental factors is one critical approach to narrowing both socioeconomic and racial gaps.

The Role of Interventions

Though we do not know the specific mechanism by which socioeconomic status affects child outcomes,
scholars commonly posit two theoretical models (Conger, Conger, & Martin, 2010). The “investment model” focuses on the direct role of income as a means for providing children the resources they need. In contrast, the “family stress model” focuses on the negative impacts of poverty on parents’ mental health and parenting ability, which in turn negatively influences children. The first model supports programs aimed at improving parents’ economic well-being through work incentives, income assistance, education supports, or other routes. The second model calls for programs that teach parents effective parenting techniques and reduce stress. We describe the evidence on these two types of parental interventions below, and then move on to the broader research base on the effects of interventions targeted directly towards young children.

**Family Resources**

Understanding the unique relationship between parental income, employment, or education and child outcomes is challenging because each is likely correlated with other relevant parental characteristics. A series of randomized welfare experiments conducted in the 1990s provided an opportunity to examine these relationships (Clark-Kauffman, Duncan, & Morris, 2003; Duncan, Morris, & Rodrigues, 2011; Morris, Gennetian, & Duncan, 2005; Yoshikawa, Gassman-Pines, Morris, Gennetian, & Godfrey, 2010; Zaslow et al., 2002; Loeb, Fuller, Kagen, & Carol, 2003). Researchers assessed the impact of several employment-based welfare programs targeted at single-parent families and found that overall the programs did not strongly affect child outcomes. However, programs that included earning supplements yielded systematic positive effects on the cognitive outcomes of children two to four years after program entry. Adolescents, however, did not benefit, and, in fact, there was some evidence that program participation was associated with detrimental outcomes, such as more frequent drug use and behavior problems in school.

Duncan and Magnuson (2004) posit that indirect interventions are likely to yield less meaningful impacts on child outcomes than are direct interventions. As an example, many policymakers have designed interventions aimed at increasing maternal education levels in the hopes that the higher education levels would lead to positive impacts on children. In order for such interventions to yield the desired results, two sequential processes must occur. The intervention would have to successfully raise maternal education, and then that maternal education would have to affect child outcomes through specific skills learned, increased earnings, or another mechanism. Experimental studies often demonstrate that even when these types of interventions do raise maternal education, the increase fails to translate into significant benefits for children (Quint, Bos, & Polit, 1997; Zaslow et al., 2002). That said, some recent quasi-experimental research shows that exogenous variation in family income created by the Earned Income Tax Credit eligibility predicts children’s health and cognitive outcomes (Dahl & Lochner, 2012; Hoynes, Miller, & Simon, 2012).

**Family Processes**

A second set of interventions aims to change parenting practices. A strong association between early literacy practices and child outcomes has led to a variety of parental interventions focused on home reading practices. One strategy involves promoting literacy during children’s regularly scheduled doctor’s visits. The Reach Out and Read (ROR) program, for instance, involves reading to children in the waiting room of clinics, giving them developmentally appropriate books to take home, and having pediatricians discuss the benefits of reading during the actual appointment. Mendelsohn et al. (2001) compared the vocabulary development of children in two similar inner-city pediatric clinics, one of which had used the ROR for three years, while the other had only recently introduced the intervention. Multivariate regressions showed that those children who had been visiting the clinic that offered ROR had significantly higher scores on measures of both receptive and expressive language. More generally, these relatively inexpensive interventions have been associated with improvements in low-income, black, and Hispanic parents’ literacy practices and children’s early language development (Blom-Hoffman, O’Neil-Pirozzi, Volpe, Cutting, & Bissinger, 2007; Golova, Alario, Vivier, Rodriguez, & High, 1999; Needlman, Toker, Dreyer, Klass, & Mendelsohn, 2005; Sharif, Rieber, & Ozuah, 2002; Zuckerman, 2009).

Parental literacy interventions carried out through child care centers and elementary schools also are
associated with improved short-term literacy outcomes for children (Jordan, Snow, & Porche, 2000; Whitehurst et al., 1994). Recent meta-analysis of family-based literacy interventions found that these programs are associated with small but significant short-term gains for young children (Manz, Hughes, Barnabas, Bracaliello, & Ginsburg-Block, 2010; Reese, Sparks, & Leyva, 2010). To our knowledge, no experimental studies demonstrate long-term benefits associated with parent literacy interventions.

Some long-term impacts emerge from more broadly aimed parenting interventions. One of the most rigorously studied examples is the Nurse Home Visitation Program, which began in a semirural area of New York in 1978 and followed families longitudinally through adolescence (Olds et al., 1997). Five hundred pregnant women were randomly placed in either a treatment or a control group. Women in the treatment group received regular home visits that focused on maternal health, positive parenting, and personal development. The control group received developmental screenings for their children at 12 and 24 months of age, but no further maternal support services. The 15-year follow-up showed that program participants were less likely to be perpetrators of child abuse and neglect. The subgroup of mothers who were unmarried and low-income at the time of their pregnancy also experienced fewer subsequent births, fewer months on welfare, and fewer arrests, and their children displayed lower levels of arrests, lower incidents of running away, fewer sexual partners, and less regular alcohol consumption. A more recent replication study among an urban sample also showed significant benefits for children (Olds et al., 2010).

While the results from this program are encouraging, findings from the growing body of research on home visitation more broadly are mixed and arguably underwhelming (Astuto & Allen, 2009; Azzi-Lessing, 2011; Gomby, 2005). While some home visitation programs result in various improvements in parenting practices, they often yield few measured effects on children’s development, and authors of review studies generally considered the few small positive effects identified as unsystematic exceptions. While early attrition and inconsistent participation may partially explain the lack of positive outcomes, the intensity of these home visitation programs is likely the most relevant factor. As Gomby et al. (1999) point out, it may be unrealistic to expect programs involving 20–40 hours of direct contact over several years to have such significant impacts on parental behaviors that children’s outcomes are affected in a meaningful and significant fashion. Nevertheless, a recent RAND report finds that in-home visitation programs yield a $2.24 return for every dollar invested, and in recent years these programs have received heightened policy attention at both the state and the federal level (Karoly, Kil-burn, & Cannon, 2006). For instance, as of 2009, 40 states support home visitation programs, and the bulk of these also relied on federal funding sources (Boonstra, 2009). The federal Affordable Care Act also allocated $1.5 billion over five years to support states’ efforts to deliver home visitation programs.

Centers

Center-based early childhood programs are among the most highly utilized direct child interventions. In 2012, over half of three-to-five-year-old children not yet enrolled in kindergarten were enrolled in some form of center-based care, and among four-year-olds over two thirds of children participate (Bassok, 2010; Child Trends, 2014; Snyder & Dillow, 2012).

Experimental Evidence. Some of the most compelling evidence on centers comes from a handful of well-designed randomized experiments. The primary advantage of experimental studies is that, when done well, they eliminate doubts about selection bias. Since children are assigned to an intervention or control group by chance, we can be fairly certain that their post-intervention results represent the causal relationship between the intervention and child outcomes. The most famous of these studies is the High/Scope Perry Preschool, described earlier, which shows significant differences between the treatment and control group at every wave of the study from preschool to age 40.

A similar experiment—the Carolina Abecedarian project—randomly assigned 111 infants born to extremely poor, high-risk mothers into a treatment and control group (Campbell et al., 2012). Both the treatment and the control group received some enriched social services; but starting at age six months, children in the treatment group were enrolled in an intensive, high-quality preschool program that ran full-day, full-year through kindergarten entry. While the cognitive performance of the two groups was roughly similar until age nine months, by 18 months a gap rapidly emerged, and a dramatic 14-point difference on an IQ test persisted through the preschool years. In the most recent follow-up, the Abecedarian children are 30, and the treatment...
group outperformed the control group substantially on educational attainment. For instance, the treatment group was far more likely to have earned a bachelor’s degree (23 percent compared to 6 percent). The treatment group was also more likely to be employed full-time, though there were no systematic differences with respect to income or a variety of social outcomes.

Taken together, the Perry and Abecedarian preschool experiments show that intensive, high-quality preschool interventions can have powerful and lasting effects, at least for very poor children. That said, both interventions targeted extremely at-risk populations, making it difficult to generalize their results to the population at large. Further, both of these programs offered high-quality care and in turn were far more expensive and intensive than more typical preschool programs. Finally, these interventions occurred decades ago, when participation in center-based care was atypical. These findings may not generalize to current programs that, for instance, may last only one year, meet only half day, or not necessarily employ highly educated teachers with engaging curricula.

The Head Start Impact Study is the most notable example of a current-day experiment to evaluate the impacts of a large-scale early childhood program. Head Start is a federally funded comprehensive early childhood program started in 1965 as part of Lyndon Johnson’s War on Poverty. Today, Head Start serves over a million families nationwide. In 2002, roughly 5,000 children eligible for Head Start were randomly assigned either to Head Start or to a control condition. While in Head Start, the treatment group outperformed the control on a number of outcomes. However, many of the gains realized by participants by the end of the program year eroded by kindergarten or first grade, and by third grade the treatment and control groups were indistinguishable (Puma et al., 2012). The lack of Head Start program effects by the early elementary years has raised concerns about the likely benefits of large-scale early interventions. Researchers and policymakers have posited a number of hypotheses to explain the lack of congruence between these findings and the early experimental evidence from small model programs including differences in program quality and intensity and differences in the experiences of the control group samples.

Nonexperimental Evidence. Although experimental studies offer the most methodologically rigorous evidence on the effects of center-based care, a growing body of quasi-experimental and descriptive research has vastly expanded our understanding of the relationship between scaled-up center-based programs and child outcomes. While determining causality is more difficult in nonexperimental analyses, these studies tend to find a positive relationship between center participation and children’s cognitive outcomes, though, as might be expected, the magnitude of the effects are smaller than those observed in the more intensive preschool experiments (Barnett, 2008; Blau & Currie, 2006; Currie, 2001). Below we summarize the evidence from three types of nonexperimental studies: (1) longitudinal studies assessing the long-term impact of preschool interventions; (2) descriptive shorter-term studies using more recent, nationally representative data; and (3) evaluations of large public preschool programs.

The longitudinal analyses of Chicago’s Child Parent Centers, which opened in 1967, provide some of the most compelling nonexperimental, long-term evidence about the impact of early childhood programs. These centers offered comprehensive social, health, and educational support services for impoverished children ages three through nine and their parents. Specifically, the program included a high-quality, half-day, nine-month preschool program, home visitation, outreach services, as well as comprehensive school-age services (reduced class size, enrichment activities, etc.) for a subsample of participants. Reynolds et al. (2011) compared the outcomes of program participants to those of demographically similar children living in the neighborhoods where centers were operating. They found that the children who were enrolled in the child parent centers were less likely to be retained or placed in special education and were more likely to graduate from high school. At age 26, program participants had lower arrest rates, lower rates of depressive symptoms, and higher rates of insurance coverage. The authors estimate the program had a return of nearly 11 dollars for every dollar invested.

Ludwig and Miller (2007) creatively apply a regression discontinuity design to provide more evidence of the positive long-term impact of early interventions from the 1960s. When Head Start was first introduced in 1965, the Office of Economic Opportunity tried to encourage take-up in high-poverty counties by providing the 300 poorest districts in the country with intensive technical assistance for funding proposals. This policy created a sharp discontinuity in program funding between those counties on either side of the technical assistance cutoff point. The authors leverage this exogenous discontinuity to assess whether counties that fell just below the
cutoff for technical assistance varied systematically on health and educational outcomes from counties that fell just above the cutoff. They report significantly lower mortality rates in those counties just below the cutoff line and demonstrate these drops in mortality are driven by declines in deaths from "Head Start susceptible causes" such as smallpox, polio, and measles, rather than injuries. They also find slightly weaker evidence of differences in educational attainment. Using within-family comparisons Deming (2009) also finds evidence that Head Start participation positively impacts young adult outcomes. Taken together, the research on both the Child Parent Studies and on Head Start participation several decades ago is consistent with the early experimental evidence in suggesting that targeted interventions several decades ago yielded substantial long-term benefits.

To explore the relationship between center-care participation and child outcomes for more recent cohorts of children a number of studies have analyzed various waves of the nationally representative Early Childhood Longitudinal Study (Bassok, Gibbs, & Latham, 2013; Bassok, 2010; Loeb, Bridges, Bassok, Fuller, & Rumberger, 2007; Magnuson, Ruhm, & Waldfogel, 2007a). All reported a positive relationship between participating in center-based care in the year prior to kindergarten and children's cognitive outcomes. Controlling for a rich set of environmental variables, Magnuson, Ruhm, and Waldfogel (2007a) found that center attendees outperformed children who stayed at home on both a reading and math assessment, with effect sizes of 0.12 and 0.10. Consistent with previous studies examining this issue, the ECLS data also show that along with the cognitive benefits, participation in center-based care particularly for extended hours is associated with somewhat heightened levels of externalizing behavior, which includes fighting, arguing, and disrupting. One important advantage of the ECLS data is that its large sample size allows for meaningful analysis of subgroup differences. Magnuson et al. (2007a) found that the relationship between center care and cognitive outcomes is particularly large for very poor children and for those with low parental education. Bassok (2010) found substantially greater impacts of center participation for nonpoor blacks than for nonpoor whites. Both studies provide evidence that targeted interventions might be useful for eliminating gaps.

Researchers and policymakers are cautious when interpreting the findings from these national datasets because the positive association between center-based care and child outcomes found in nonexperimental longitudinal data may be driven by a selection mechanism, whereby those children who attend center care have unobserved characteristics that influence both their entrance into this type of care and their eventual outcomes.

A number of recent quasi-experimental studies of state preschool programs get around these selection issues in novel ways (Fitzpatrick, 2008; Weiland & Yoshikawa, 2013; Wong, Cook, Barnett, & Jung, 2008). Gormley et al. (2005), for instance, was the first of several studies to leverage birthday cutoffs for public preschool programs as a source of exogenous variation in preschool enrollment. Their study evaluated the effects of Oklahoma’s universal preschool program, which established a free, voluntary preschool, run through public schools and open to all four-year-olds in the state. Almost three-quarters of the eligible children in Oklahoma participate, which is among the highest penetration rate nationwide. To minimize selection issues, Gormley et al. used a regression discontinuity design, comparing a treatment group comprised of children who enrolled in one year, and a control group of those children whose birthdays were after the cutoff date but who enrolled in the following year. Adjusting their regressions for age as well as demographic characteristics, the authors found a 0.79 of a standard deviation gap on a letter-word identification measure and a 0.38 standard deviation gap on an applied problem measure between the treatment and control group. Further, they showed that the positive pre-school effects arise in all racial and socioeconomic subgroups, though they were unable to compare the magnitude of the effects across groups.

**Sustaining Effects.** The evidence from large nationally representative datasets as well as from recent evaluations of state preschool programs tends to show that participation in center-based care is associated with a variety of cognitive benefits at school entry. Some recent studies also suggest positive medium-term benefits from state early childhood initiatives in Oklahoma, Georgia, and North Carolina (Fitzpatrick, 2008; Hill, Gormley, & Adelstein, 2012; Ladd, Muschkin, & Dodge, 2014). However, as with the findings from the Head Start Impact Study discussed above, many recent studies suggest that the benefits gained from preschool participation are often short-lived, dissipating early in elementary school (Karoly et al., 1998; Magnuson, Ruhm, & Waldfogel, 2007b). For instance, a recent experiment evaluating the impact of Tennessee’s Voluntary Pre-K Program showed substantial impacts at the end of the preschool year, but these effects diminished.
rapidly and were no longer statistically significant by the end of the kindergarten year (Lipsey, Hofer, Dong, Farran, & Bilbrey, 2013).

Researchers and policymakers are eager to understand why the effects of programs such as the Perry Preschool, Abecedarian, and the Child Parent Centers last into early adulthood and beyond, and that even the Head Start programs from several decades ago demonstrate long-lasting effects, while the benefits of more recent and more broad-based preschool programs seem to fade out after the first few years of school. At least four interrelated forces may explain this fade-out: program quality, program intensity, school experiences in the years after preschool, and noncognitive skills as the driving mechanism for long-term gains.

A limitation of much of the data used to study the effects of preschool is the lack of information on the quality of centers that children attended. Early childhood programs in the United States face widely varying regulations, and quality varies substantially across settings (Bassok, Fitzpatrick, Greenberg, & Loeb, 2014). While the results we describe above characterize the average effects of centers, we do not know whether these effects vary based on characteristics of curriculum, instruction, or facilities. Magnuson et al. (2007a) attempt to address this issue using the ECLS-K data by separately analyzing program effects for children whose parents indicated they attended a “pre-kindergarten” program and those who attended other types of centers (e.g., nursery school, daycare, or preschool). The authors assume that the pre-kindergarten programs are school-based programs that have higher levels of regulation and a more educational curriculum. Indeed, they find that cognitive outcomes are higher in the pre-kindergarten group, though levels of externalizing also appear higher in these settings.

A large body of literature has more directly examined the effects of observable measures of care quality on child outcomes (Loeb, Fuller, Kagan, & Carrol, 2004; Mashburn et al., 2008; NICHD ECCRN, 2003; Feisner-Feinberg et al., 2001). These studies all indicate that higher-quality centers in terms of both structural and process measures have larger and more long-lasting effects on children’s cognitive and social outcomes. Aggregating the effects of all preschool programs together may result in low-quality programs masking long-lasting effects from other higher-quality programs. Indeed, many states as well as the federal government are now investing in Quality Rating and Improvement Systems as a way to track program quality and incentivize a focus on quality enhancements.

Another possible explanation for the fade-out of preschool effects stems from the intensity or dosage of early childhood interventions. Children’s exposure to center-based programs varies dramatically in terms of both age at entry and hours per week. To the extent that more intensive interventions are stronger predictors of long-term benefits for children, aggregating the outcomes of all center attendees irrespective of the amount of time they have spent in center could mask effects (Loeb et al., 2007). A number of recent papers have compared the impacts of attending preschool for two years rather than one, and all have suggested that more exposure is beneficial (Arteaga, Humpage, Reynolds, & Temple, 2013; Domitrovich et al., 2013; Wen, Leow, Hahs-Vaughn, Korfmacher, & Marcus, 2011). Evidence from the Chicago Child Parent Centers further supports this notion that amount of exposure matters, because those children who extended their participation into the first years of elementary school displayed higher levels of school achievement from elementary school through high school, and also had significantly fewer special education placements (Reynolds, Temple, Robertson, & Mann, 2001).

A related third explanation for the fade-out of child care interventions has to do with the experiences children have once they enter school. Several studies have considered the possibility that school quality plays an important role in sustaining or negating the effects of center-based programs. Lee and Loeb (1995) test this hypothesis within the context of Head Start. Using data from the National Education Longitudinal Study of 1988, a nationally representative sample of approximately 25,000 eighth graders, the authors compare the school quality of children who reportedly attended Head Start, another type of preschool program, or no preschool intervention. Defining school quality broadly to include social composition, academic excellence, perceived safety, and teacher student relationships, the authors find that eighth graders who attended Head Start end up in significantly worse schools compared to children who attended other types of preschools, even when the authors control for race, parental education, and an income-to-needs ratio. They conclude that the poor quality of schools attended by Head Start participants seriously undermines the likelihood of sustained program effects.

Currie and Thomas (2000) expand on this analysis by first noting that the extent of fade-out systematically differs for blacks and whites, with the effects on blacks disappearing rapidly, but white children experiencing
long-lasting benefits through adolescence. They argue that if, indeed, the school quality explanation is driving fade-out, the difference in school quality for Head Start participants and non-Head Start participants must be larger for blacks than for whites. Their results support this claim, and they suggest that environmental factors in the years after a child completes a center-based intervention can seriously negate the initial impact of the program. In a separate study, Garces, Thomas, and Currie (2002) report that white Head Start attendees experience significant long-term gains from Head Start participation, including higher levels of high school completion and earnings during early adulthood.

This evidence suggests that the benefits of early childhood intervention would be maximized if combined with high-quality experiences for participants as they enter elementary schools. In fact, the Race to the Top Early Learning Challenge (RTT-ELC), a federal grant competition, identified "creating preschool through third grade approaches to sustain improved early learning outcomes through the elementary grades" as a key component for a competitive application.

Finally, Gibbs et al. (2013) posit that the observed fade-out of preschool effects on cognitive test scores may not be inconsistent with meaningful, long-term benefits on adult outcomes. The authors point out that a number of studies of early interventions, including even Perry Preschool, demonstrate significant long-term benefits on policy-relevant outcomes, while also showing fade-out on cognitive measures (Chetty et al., 2011; Deming, 2009; Ludwig & Miller, 2007). One explanation for this pattern may be that the positive long-term benefits of early interventions operate through impacts on noncognitive rather than cognitive domains, which are often not measured or not measured well (Cunha & Heckman, 2010).

Conclusions

This review has highlighted several crucial points about the early achievement gap. First, systematic differences in development and ability emerge long before children enter school, and those children who start school at a disadvantage are likely to remain behind their peers throughout school and beyond. Second, advances in research have highlighted the importance of the early childhood environment as a contributing factor to this school readiness gap. And third, despite our continuously expanding understanding of child development, implementing programs that meaningfully narrow the gap is a challenging task. The effects of center-based programs are highly dependent on quality and are, at times, disappointing or fleeting.

Nevertheless, increasingly there are compelling examples of interventions that confirm the positive potential of early childhood interventions to narrow achievement gaps (Duncan & Sojourner, 2013). We know that the most successful of these programs are intensive and involve substantial investments. We also know that the biggest program benefits are typically seen in the neediest children. Given the vast impediments that very poor children face, neither of these findings is particularly surprising. The bottom line is that there are no easy solutions. Achieving sizeable practical results necessarily means making a highly targeted, long-term commitment to children most in need.

References


Increasing the Effectiveness of Teachers in Low-Performing Schools

MELINDA ADNOT AND JAMES WYCKOFF

Introduction

In the United States, public education has long been considered an important tool for promoting economic opportunity, increasing social mobility, and preparing students for workforce participation. Educational attainment is strongly predictive of future labor market success, and technological advances in the U.S. economy will increasingly reinforce this relationship (Levy & Murnane, 2012). Yet, for generations poor and nonwhite students have achieved at substantially lower levels than their more affluent, white peers. Students from low-income and minority families enter school academically behind their more affluent, white peers, and there is scant evidence to suggest that these achievement gaps close as they progress through the K–12 system (Reardon, Robinson-Cimpian, & Weathers, this volume). These lower rates of achievement are associated with lower high school graduation rates, lower college attainment, increased risky behaviors, and reduced economic success (Duncan & Magnuson, 2011; Chetty, Friedman, & Rockoff, 2011). In this chapter we explore the role that teachers can play in improving outcomes for poor, nonwhite, low-achieving students.

Although many factors influence student academic achievement, recent research clearly demonstrates the importance of teachers (see Goldhaber, this volume). Effective teachers can substantially attenuate meaningful differences in student achievement (Aaronson, Barrow, & Sander, 2007; Rivkin, Hanushek, & Cain, 2005; Rockoff, 2004) and later life outcomes (Chetty, Friedman, & Rockoff, 2011). Unfortunately, large differences within and across schools in teacher quality measured both by the credentials of teachers (Clotfelter, Ladd, & Vigdor, 2006, 2007; Lankford, Loeb, & Wyckoff, 2002) and their effectiveness (Isenberg et al., 2013) typically work to the disadvantage of low-income or low-performing students.

The two main strategies for improving the quality of teachers in low-performing schools are, first, differential selection and retention of effective teachers and second, systematic improvement and utilization of existing teachers. In this chapter we examine evidence on the potential for these strategies to improve outcomes in schools serving low-income and low-performing students. We begin, however, with a summary of the context relevant for understanding the challenges confronting efforts to improve teacher performance in low-performing schools.

Context

Four features of K–12 education influence the market for teachers and are particularly relevant for understanding the challenges confronting low-performing schools: neighborhood attendance zones, teacher preferences for working conditions, salary schedules that differentiate only by experience and education, and the lack of rigorous, consequential teacher evaluation systems. Some of these features have been subject to recent change, which affords an opportunity to examine their effects.

- Most public school students attend schools defined by neighborhood attendance zones that lead to substantial sorting of students to schools due to the geographic segregation of households by income and race (Orfield, Kucsera, & Siegel-Hawley, 2012). Although various public school choice programs, including magnet and charter schools, may afford students opportunities beyond their neighborhood...
schools, these opportunities are limited and may in fact exacerbate student sorting.

- Teachers typically prefer to teach in schools with smaller percentages of poor, nonwhite and low-performing students (Boyd, Lankford, Loeb, & Wyckoff, 2005; Clotfelter, Ladd, & Vigdor, 2005 and 2011; Jackson, 2009; Lankford et al., 2002). In addition many teachers prefer to teach closer to where they grew up (Boyd et al., 2005) and in schools similar to those they attended (Ronfeldt, 2012).

- Nearly all teachers are compensated by a salary schedule that rewards teachers based primarily on experience and advanced degrees. Although some urban districts establish salary differentials intended to compensate for the challenges that arise in schools serving large proportions of low-income students, the differentials are typically too small to counter the well-documented relative shortages of effective teachers in more difficult-to-staff subjects and schools (Jacob, 2007; Ingersoll & Perda, 2009, 2010).

- Historically, the performance of most teachers has not been rigorously assessed, thereby limiting the feedback teachers receive on how to improve their teaching. As recently as 2009, most school districts evaluated teachers on a binary satisfactory/unsatisfactory outcome with more than 95 percent of teachers receiving the highest rating (Weisberg, Sexton, Mulhern, & Keeling, 2009). Lack of rigorous evaluation coupled with policies that grant tenure easily and make removing persistently ineffective teachers difficult (National Council on Teacher Quality, 2012) contributed to an environment in which excellence was not recognized, improvement was not supported, and ineffective teaching was tolerated.

These features of teacher labor markets increase the difficulty of attracting, retaining, and improving teachers generally, but especially in the lowest-performing schools. First, as students are frequently sorted by race, income, and performance and many teachers appear to prefer to work in schools with more affluent, higher-performing students, schools with higher percentages of poor and low-performing students find it difficult to recruit and retain effective teachers. As a result, schools with large shares of low-performing and low-income students have teachers with much weaker qualifications, including experience, than schools with higher-performing and/or more affluent students (Betts, Reuben, & Danenberg, 2000; Lankford et al., 2002; Hanushek, Kain, & Rivkin, 2004; Clotfelter et al., 2005, 2006, and 2011). The task of recruiting and retaining teachers with strong qualifications is made more difficult as teachers often prefer to teach close to where they attended school and in schools similar to those they attended.

Increasing evidence indicates that schools with disproportionate shares of low-performing, poor, and nonwhite students also have less effective teachers as measured by value-added, a measure of teacher quality that is based on the gains in test scores of their students (see Donaldson and Papay, this volume). In a study of 29 school districts, Isenberg et al. (2013) find that for some districts there is no difference in effective teaching between disadvantaged and non-disadvantaged students, while in other districts the gap is 0.11 student achievement standard deviations. Sass, Hannaway, Xu, & Feng (2012) find modest differences in teacher effectiveness between high- and low-poverty schools in North Carolina, but a much higher concentration of the lowest-performing teachers in high-poverty schools.1

Additionally, teachers, like employees in other occupations, appear responsive to wages and working conditions (see, e.g., Dolton & Van der Klaauw, 1995, 1999; Hanushek et al., 2004; Murnane et al., 1991; Stinebrickner, 2001). As alternative occupations became more accessible to women during the 1970s and the relative wages of teachers fell from 1960 through 1990, the academic ability of entering teachers declined (Bacolod, 2007; Corcoran, Evans, & Schwab, 2004). The interaction of features of teacher labor markets, teacher preferences, and market forces particularly disadvantaged low-performing schools. Teachers’ revealed preference for high-achieving students and constraints imposed by salary schedules on compensating differentials across schools within districts meant that low-performing schools experienced the declining interest in teaching among academically able individuals most dramatically. Because researchers lacked good measures of teacher effectiveness during this period, it is unclear whether the decrease in academic ability of teachers implies a decrease in teacher effectiveness, although recent research suggests a modest link between measures of academic ability and teaching effectiveness (Clotfelter et al., 2007; Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2008).

It is against this context that we explore the literature on strategies to increase the effectiveness of teachers in low-performing schools. We organize this review by examining evidence across two broad domains: altering
the composition of teachers in these schools through policies, such as improved recruitment and retention of effective teachers, and improving the human capital and utilization of existing teachers through policies such as teacher preparation and professional development. Before turning to these issues, we briefly consider the measurement of teacher effectiveness.

**Assessing Teacher Effectiveness**

Other chapters in this volume (see Goldhaber and Papay & Kraft) provide a more detailed understanding of teacher evaluation, but we touch on this issue to provide a foundation for the remainder of our discussion. Ideally, school administrators would have accurate measures of teacher effectiveness and would use these measures to identify teachers who are performing well and deserve recognition for their skills, teachers who would benefit from targeted professional development, and teachers who are unable to improve sufficiently to be effective and should be removed from the teaching workforce.

The debate over what constitutes effective teaching and how best to capture this multidimensional construct is hotly contested. Many states and districts now employ value-added measures that quantify teacher effectiveness by identifying the portion of gains in the standardized test scores of students attributable to the teacher. At the same time, value-added measures of teacher effectiveness have several well-identified limitations (see Donaldson & Papay, this volume). Recent progress on this front shows that combining multiple measures of performance (e.g., value-added, rigorous teacher observations and student surveys) appears to produce a far better predictor of future performance than the use of any one indicator alone (Mihaly, McCaffrey, Staiger, & Lockwood 2012). There seems to be general agreement that a measure of teacher effectiveness should address several goals: validity, reliability, formative value, cost, transparency, and face validity (Papay, 2012).

Without getting too far from the focus of this chapter, suffice it to say that having a robust teacher evaluation system is crucial for policies intending to provide support to teachers to improve their teaching practice and to selectively retain effective teachers.

**Differential Selection and Retention of Effective Teachers**

Differential selection and retention includes the recruitment of effective novice teachers, the attraction and retention of effective veteran teachers, and the exit of less effective teachers. Each has the potential to influence the overall effectiveness of the teacher workforce. Policymakers, stakeholders and researchers have invested heavily over the last decade to improve the quality of the teacher applicant pool and to identify candidates who will become effective teachers. At the federal level, the No Child Left Behind Highly Qualified Teacher provision increased required teacher qualifications. Many school districts and stakeholder organizations, such as Teach for America (TFA) and TNTP (formerly The New Teacher Project) focused efforts on identifying attributes of entering teachers that predict later effectiveness. Differential selection and retention among current teachers has received increasing attention as measures of teacher effectiveness have become more common, allowing policymakers to address teacher effectiveness directly.

**Attracting Effective Novice Teachers to Low-Performing Schools**

 Recruiting novice teachers to low-performing schools has been a long-standing challenge. If schools wish to hire teachers who will ultimately become effective at improving student learning (Boyd, Lankford, Loeb, & Wyckoff, 2013), they place a premium on identifying attributes of novice teachers that are strong signals of their future effectiveness. Unfortunately, easily observed attributes of pre-service teachers do not predict on-the-job performance very well (Rice, 2003; Rivkin et al., 2005). Traditional credentials such as academic background (Harris & Sass, 2011), certification exam scores and certification status (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2006; Kane, Rockoff, & Staiger, 2008), masters degrees (Clotfelter et al., 2007) and college
entrance exam scores (Harris & Sass, 2011) individually provide weak signals of future productivity. However, when taken together, these attributes provide a stronger, but still modest, signal of teacher ability (e.g., with an effect size of 0.04) to improve student achievement (Boyd, Lankford et al., 2008; Clotfelter et al., 2007).

Recent studies have moved beyond the observable teacher characteristics available in administrative datasets to explore how personal traits such as leadership and personality may predict future effectiveness (Rockoff & Speroni, 2010; Rockoff, Jacob, Kane, & Staiger, 2011; Dobbie, 2011; Duckworth, Quinn, & Seligman, 2009). While some of these efforts appear to produce relatively large effect sizes, e.g., Dobbie’s examination of TFA selection criteria (0.15), many are associated with modest gains in achievement.

Even if one can identify novice teachers who are likely to be effective, finding mechanisms to attract them to low-performing schools has been difficult. Financial incentives, if sufficiently large, can encourage new teachers to accept positions in low-performing schools. For example, Steele, Murnane, and Willet (2010) find in California that from 2000 to 2002 receipt of a $20,000 Governor’s Teaching Fellowship increased the probability of a high-ability candidate entering a low-performing school by 28 percentage points. However, most policies designed to attract teachers include relatively small incentives for which there is little evidence of success (Goldhaber, 2008). Student loan forgiveness has been used as an incentive to attract new teachers to low-income schools, but there is little rigorous evidence on its ability to increase teacher quality.

Alternative routes to certification are another approach to expanding the pool of teacher applicants and have become an especially important source of teachers for low-performing schools that often faced chronic shortages of traditionally prepared teachers. Several experiments (Glazerman, Mayer, & Decker, 2006; Constantine et al., 2009; Clark et al., 2013) and rigorous quasi-experimental studies (Boyd et al., 2006, 2012; Kane et al., 2008, Xu, Hannaway, & Taylor, 2011) have compared the effectiveness of alternatively certified teachers to traditionally certified teachers. There is ample evidence that highly selective alternative certification routes, e.g., TFA or the New York City Teaching Fellows, can produce teachers whose students perform as well or better than students of traditionally certified teachers in math (Glazerman et al., 2006; Boyd et al., 2006, 2012; Clark et al., 2013; Constantine et al., 2009, Kane et al., 2008, Xu et al., 2011). The evidence on reading is mixed. In some cases TFA teachers are no different than traditionally prepared teachers (Glazerman et al., 2006); in other cases they are less effective (Boyd et al., 2006). Generally, students of Teaching Fellows perform less well in reading than teachers from traditional preparation programs (Boyd et al., 2006, Kane et al., 2008).

In a recent experimental study, alternatively certified teachers from less-selective programs were found to be neither more nor less effective than traditionally certified teachers (Constantine et al., 2009). However, this study was designed to have a minimum detectable effect size greater than 0.10 SD, so given that many studies identify differences of 0.02 to 0.07 SD, the null effect found in this study does not distinguish it from a Florida study finding variability between teachers from different nonselective alternative routes (Sass, 2011). Sass finds that teachers from the most successful program, American Board for Certification of Teacher Excellence (ACBTE), outperform their traditionally certified peers by 0.06 to 0.11 SD in math, while teachers from education preparation institutes, the least successful programs, lag by about 0.03 to 0.04 SD.

Although teachers from highly selective alternative certification programs perform no worse and perhaps better than teachers from other routes, many are less likely to persist, especially TFA teachers (Darling-Hammond, Chung, & Frelow, 2002; Heilig & Jez, 2010; Boyd et al., 2006, 2012). However, after accounting for the higher rates of turnover in high-need schools their performance appears, on average, to be no worse than that of teachers from other routes (Boyd et al., 2006, 2012; Xu et al., 2011).

In sum, identifying effective teachers prior to hire is conceptually appealing as a means of promoting higher student outcomes, minimizing costs, and promoting a strong, cohesive culture of learning. However, recent research indicates a modest ability to employ observable teacher attributes to distinguish effectiveness a priori. As a result, many individuals who become teachers may not have strong teaching abilities on entry. As the least effective teachers are more likely to teach in low-performing schools, it would be useful to pursue policies to redistribute effective teachers, to improve the effectiveness of teachers, or to remove the least effective teachers.

**Incenting Transfers for Effective Teachers**

456
Given the difficulty of predicting future effectiveness at the time of hire, attracting veteran teachers who have already demonstrated their effectiveness could be a promising strategy for raising the performance of teachers in schools with low-performing students. Empirical work suggests that the compensating differential necessary to attract and retain effective teachers to schools with high percentages of low-income or low-performing students is large (Clotfelter et al., 2011; Boyd et al., 2013). An alternative, which has only recently been tried, is to design, financial incentives to encourage strong performers to move to more difficult-to-staff schools.

One such effort, The Talent Transfer Initiative (TTI), employed a randomized field trial to assess whether high-performing teachers would transfer and remain effective in low-performing schools (Glazerman, Protik, Teh, Bruch, & Max, 2013). TTI offered teachers from the top 20 percent of the value-added distribution in each of 10 school districts $20,000 to transfer to a low-achieving school in their district and remain there for two years. Among eligible receiving schools with comparable grade and subject level vacancies TTI randomly assigned teachers to a treatment vacancy (eligible for the TTI incentives) or a business-as-usual control vacancy.

Several noteworthy findings emerge from the study. First, only 22 percent of teachers identified as high-performing and contacted completed an online application to transfer, with just 5 percent ultimately transferring. Second, high-performing elementary treatment teachers outperformed control teachers in both the year following transfer and the next year in reading and math with differences ranging from 0.10 to 0.25 SD depending on year and subject. In middle school the differences were generally much smaller and usually not statistically significant. Finally, TTI teachers had higher retention rates than control teachers following the first year of the study, when treatment teachers were still receiving payments. Notably, following the second year and the conclusion of TTI payments, retention differences between TTI and control teachers were not statistically significant.

In sum, the TTI suggests that enticing highly effective teachers to transfer to low-performing schools is feasible, and these teachers appear to continue to be high-performing in their new schools. However, relatively few teachers appear interested in transferring, and, for those who are, financial incentives need to be large and remain in place suggesting the need for some form of performance-based pay differentials, not just one-time transfer bonuses.

Teacher Attrition in Low-Performing Schools

Prevailing wisdom holds that early-career teachers attrit at high rates, and especially so in schools with disproportionate numbers of low-achieving, nonwhite, and poor students (Boyd et al., 2005; Darling-Hammond & Sykes, 2003; Hanushek et al., 2004; Ingersoll, 2001). However, recent comparative work indicates that teacher turnover is no higher for early-career teachers than for novices in comparable professions such as nursing and social work (Harris & Adams, 2007). So while some attrition might be under the control of schools or districts, some attrition should also be expected as a result of unavoidable personal decisions and increasingly customary early-career transitions.

Nonetheless, major school staffing problems result from the “revolving door,” whereby large numbers of qualified teachers transfer or exit the profession (Ingersoll, 2001). High attrition rates can be disruptive to school cohesion and climate (Bryk & Schneider, 2002; Hanselman, Grigg, Bruch, & Gamoran, 2011) and place a burden on school resources (Barnes, Crowe, & Schaefer, 2007). Recent quasi-experimental work finds that teacher turnover from schools disrupts student achievement independently of the effectiveness of departing teachers, especially in schools with large proportions of low-performing and black students (Ronfeldt, Loeb, & Wyckoff, 2013), as does churning within schools as teachers move to teach in different grades or subjects (Atteberry, Loeb, & Wyckoff, 2014).

At the same time, a growing body of research finds that less effective teachers are more likely to leave—especially early in their careers—than their more effective peers (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2008; Boyd, Lankford, Loeb, Ronfeldt, & Wyckoff, 2011; Goldhaber, Gross, & Player, 2011; Hanushek & Rivkin, 2010). As a result, some attrition is potentially beneficial for schools. Moreover, Hanushek and Rivkin (2010) observe that the difference in effectiveness between stayers and leavers is even larger in schools with a large number of black and low-achieving students, suggesting that targeting retention policies strategically to more effective teachers and allowing or encouraging less effective teachers to leave may be particularly relevant in
high-need schools.

Taken together, this research suggests that some teacher attrition from low-performing schools is to be expected, and can be beneficial in that low-performing teachers are more likely to voluntarily exit. Nonetheless, turnover per se exacts a cost on achievement. Additionally, many effective teachers transfer or exit, especially in low-performing schools, compromising student achievement.

**Retaining Effective Teachers**

While the consequence of losing the highest-performing teachers is clear, we know surprisingly little about how to retain these teachers. Interestingly, in a survey of teachers in four districts, high-performing teachers reported that little effort was made to encourage them to stay—only 37 percent were encouraged by their principal to remain at their school the next year (TNTP, 2012).

Teachers are sensitive to working conditions, which include student, administrator, and school climate factors (Boyd et al., 2003; Jackson, 2009). Some of these factors are malleable, such as school leadership, while others are less so, such as the racial and economic composition of the students. Researchers employing teacher surveys in California, New York, Massachusetts and North Carolina examine how teachers’ perceptions of their working conditions affect retention (Boyd, Grossman, Ing, Lankford, & Wyckoff, 2011; Johnson, Kraft, & Papay, 2012; Ladd, 2011; Loeb, Darling-Hammond, & Luczak, 2005). Collectively, they find that school leadership typically emerges as the most salient factor, while collegial relationships among staff and resource considerations are also significant predictors of teacher turnover after controlling for student characteristics. Teacher influence over policy decisions (Boyd et al., 2011) and classroom autonomy (Grisom, 2011; Ingersoll, & May, 2012) also predict teacher retention. School leaders may strategically assign teachers that they wish to retain to more desirable assignments to increase retention (Loeb, Kalogrides, & Béteille, 2012).

Three recent papers explore the effects of financial incentives on teacher retention. First, an $1,800 retention payment made to math, science, and special education teachers in North Carolina secondary schools with large percentages of poor or low-achieving students reduced attrition by an average of 17 percent (Clotfelter, Glennie, Ladd, & Vigdor, 2008). Although the authors did not directly examine how the policy affected teacher effectiveness, the policy impact was greatest among more experienced teachers, which, given returns to experience, suggests it may well have improved teacher quality in the affected schools.

Research examining the effects of the Teacher Advancement Program (TAP), which includes performance-based compensation, professional development and career advancement, finds that TAP improved retention (Glazerman & Seifullah, 2012). Piloted in Chicago from 2007–08 to 2010–11, the Chicago TAP program awarded bonuses based on classroom observations and school- or grade-level value-added that averaged between $1,100 and $2,500 over the four years of the program. Teachers could earn much larger incentives if selected as mentor or master teachers. Teachers working in TAP schools in the first year of the study were 20 percent more likely than non-TAP teachers to remain three years later. The study is unable to discern whether the incentives improved retention differentially for more effective teachers.

Finally, research employing a regression discontinuity design to examine the effects of IMPACT, the District of Columbia Public Schools teacher evaluation system, finds that very strong financial incentives targeted to teachers identified as among the highest-performing teachers increases retention but the finding is not statistically significant (Dee & Wyckoff, 2013). Retention among these teachers is already high; fewer than 8 percent of highly effective teachers exited the district.

In sum, retaining effective teachers in low-performing schools remains an important goal. While evidence suggests that financial incentives may be a useful policy tool, a great deal of descriptive evidence suggests that other working conditions, including school leadership, the racial and economic attributes of students, and collegiality among teachers are important in teacher retention decisions (for example, see Clotfelter et al., 2011). More evidence on policies that would address these issues is needed.

**Removing Ineffective Teachers**

Teachers are rarely held to rigorous evaluation standards and even less often dismissed for poor performance
Given research that poor, low-performing schools are disproportionately likely to have ineffective teachers, the effects of improving teacher skills (explored below) or dismissing ineffective teachers are particularly salient to these schools. This issue has been explored both by simulation and quite recently by quasi-experimental designs.

Staiger and Rockoff (2010) develop a simulation model of optimal retention of novice teachers based on assumptions of little to no prehire signal of effectiveness and a noisy performance signal in the early career. Based on the simulation, the authors conclude that teacher effectiveness is optimized by dismissing between 70 and 80 percent of novice teachers. In a similar vein, Hanushek (2009) estimates that eliminating the bottom 6–10 percent of the workforce overall would result in a 0.5 SD increase in student achievement. In response, Rothstein (2012) argues these models should incorporate a dynamic labor market supply, recognizing that the supply of teaching entrants is limited and will be sensitive to the increased risk of higher dismissal rates.

In 2009–10 the District of Columbia Public Schools (DCPS) implemented IMPACT, a teacher evaluation system that stipulates that teachers rated "ineffective" must be immediately dismissed and those rated "minimally effective" are afforded a year to become "effective" or they will be dismissed. Every teacher is rated and all evaluations are based on multiple measures of teacher effectiveness. During its first three years IMPACT led to the performance-based dismissals of about 340 teachers (3.2 percent of all teachers). Additionally, Dee and Wyckoff (2013) find the incentives embedded in IMPACT induced many more poorly performing teachers to voluntarily exit the district. Employing a regression discontinuity design, they find "minimally effective" teachers, who would be dismissed if their performance did not improve sufficiently to be rated "effective" by the next year, were zero percent more likely to voluntarily exit the district than otherwise similar teachers who just surpassed the "effectiveness" threshold and thus were not under the threat of dismissal. Due to provisions in the DCPS teacher bargaining agreement, the district was able to implement IMPACT without union approval. That provision is rare and may mean that other districts are unlikely to implement teacher evaluation systems with equivalent use of teacher dismissals. The authors address concerns about teacher supply by showing that entering teachers were rated substantially higher than the average of all exiting teachers. Labor supply conditions may not be as favorable in other districts.

Nearly all states have some form of a tenure review process, which includes a probationary period during which teachers are intended to demonstrate they have the requisite teaching and professional conduct skills to warrant tenure. Yet, the receipt of tenure is an expectation for nearly all teachers and traditionally has involved only a superficial evaluation following a probationary period typically of three or fewer years (National Council on Teacher Quality, 2012). Following the receipt of tenure, most states grant teachers strong job protections in the form of due process, which make dismissals difficult and time consuming. However, many states and school districts are now infusing the tenure process with more rigorous data on teacher performance (National Council on Teacher Quality, 2012).

As an example, beginning in 2009–10, New York City initiated a reform in the way it approached tenure, such that fewer than 60 percent of teachers were granted tenure post-reform, far below the 94 percent before reform (Loeb, Miller, & Wyckoff, 2014). Most teachers who were not granted tenure post-reform had their probationary periods extended for a year. The authors show that attrition among these "extended" teachers was twice as great as otherwise similar teachers who were approved for tenure. Additionally, the extended teachers who exited were less effective than those who persisted. It is too early to determine the causal effects of this reform on overall teacher effectiveness and student achievement, but increasing the rigor of tenure review is an option that is available in most school districts.

Recent work has also examined the effect of a selective involuntary teacher transfer policy on teacher effectiveness and student achievement in low-performing schools (Grissom, Loeb, & Nakashima, 2013). For three years Miami-Dade County Public Schools invoked their ability to transfer teachers, prioritizing the teachers that low-achieving schools identified to transfer to other schools. Grissom, Loeb, and Nakashima (2013) find that principals from these schools identified teachers with lower value-added and more absences for transfer, and were able to hire teachers with higher value-added and fewer absences, increasing overall teacher effectiveness in these schools. While involuntary transfer policies are often disallowed by collective bargaining agreements, this analysis suggests they could be effective in more equitably distributing effective teachers.

The research on removing ineffective teachers is limited but evolving quickly. We need to know much more about policies of this sort, including the ability to implement them more broadly, their impact on recruitment,

(Weisberg et al., 2009).
and their effects on student achievement.

Systematic Improvements in Teacher Job Performance

Improving the job performance of existing teachers in low-performing schools is appealing on many levels; not least among these is its ability to redress weaknesses among teachers which could not be screened pre-service and minimizing the reliance on dismissals. Improvements in teacher performance may result from a variety of experiences, including pre-service training, in-service professional development, and less formal on-the-job learning.

Pre-Service Teacher Preparation

Pre-service teacher training offers an opportunity to provide teachers with the knowledge and skills that will allow them to be successful in low-performing schools. An increasing number of studies employ measures based on the gains in learning of the students taught by graduates to measure the effectiveness of teacher preparation programs (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Harris & Sass, 2011; Noell, Porter, Patt, & Dahir, 2008). While potentially useful for accountability purposes, providing a single estimate of program effectiveness is not very insightful for program design or improvement. The evidence on what attributes of pre-service training are most effective is very limited. Boyd et al. (2009) employ rich administrative data augmented with a survey of teachers and detailed descriptions of 18 teacher preparation programs to examine which components of preparation were associated with student achievement. They find that teachers who engaged in the actual work of teaching, e.g., greater engagement in and supervision of student teaching experiences, during preparation have students who demonstrate greater learning. In contrast, Harris and Sass (2011) find no effect of pre-service training. Their analysis was limited to the use of undergraduate transcript data and thus was unable to detect differences in content or quality beyond that signaled by course titles.

Ronfeldt (2012) explores the role of learning context in the pre-service student teaching experience. He finds that teachers who had their student teaching placement in schools where teachers experienced less attrition, had students with larger achievement gains, even when these teachers eventually taught in the most difficult-to-staff schools with the lowest-performing students. This suggests that the easier-to-staff schools have developed the capacity to provide teachers with human capital that generalizes across contexts. These findings complement work by Jackson and Elias (2009) who find that about 20 percent of a teacher’s value-added student achievement gains are attributable to their prior colleagues.

Recent evidence suggests that the decline in the academic ability of entering teachers referenced earlier has reversed and that it has been increasing since about 2000. Based on 30 years of data for SAT takers in New York State, Lankford, Loeb, McEachin, Miller, and Wyckoff (2014) document increasing trends in the test scores of those becoming teachers, including those in low-performing schools. Goldhaber and Walch (2014) employ the Baccalaureate and Beyond data to find improvements in academic ability of individuals opting to become teachers nationally. At this point, it is unclear what caused these improvements but they are consistent with the increased policy attention on teacher qualifications and quality.

In-Service Professional Development

Despite large investments in professional development, researchers have been unable to identify programs that are consistently effective at-scale (Wayne, Yoon, Zhu, Cronen, & Garet, 2008). However, the pooled effect size across the small-scale, developer-delivered studies in a recent What Works Clearinghouse review was remarkable (0.55 student achievement SD) and provides proof of the efficacy for professional learning methods used in the trials (Yoon, Duncan, Wen-Yu Lee, Scarlloss, & Shapley, 2007).

More recently, researchers and policymakers have begun to explore a new model of targeted professional
support that assesses teachers on their use of evidence-supported classroom practices and provides targeted feedback and coaching where weaknesses are observed (Allen, Pianta, Gregory, Mikami, & Lun, 2011; Taylor & Tyler, 2012). While more work is needed in this area, early results have been encouraging.

On-the-Job Learning

On average, novice teachers improve in effectiveness as they gain experience in their early careers (Atteberry, Loeb, & Wyckoff, 2013; Boyd, Lankford et al., 2008; Harris & Sass, 2011; Papay & Kraft, 2013). This early career productivity growth is not surprising; theory from a variety of fields indicates that workers will build "job-specific human capital" or engage in "informal learning" with the greatest intensity during their first few years on the job (Becker, 1993; Eraut, 2004). Through practice, observation, and formal learning opportunities, novice teachers have numerous opportunities to become more effective during their first few years of teaching.

The steep growth in the average performance of novice teachers is characterized by substantial variability, and this variability offers an opportunity to better understand the determinants of improvement (Atteberry et al., 2013; Kraft & Papay, in press). It is estimated that 30 percent of this variation among teachers results from differences across schools, while the remainder reflects differences between teachers within schools (Atteberry et al., 2013). Indeed, more supportive professional environments (Bryk, Sebring, Allensworth, Easton, & Luppescu, 2010; Kraft & Papay, in press), more effective schools and peers (Jackson & Bruegmann, 2009; Loeb, Kalogrides, & Béteille, 2012), and fewer high-poverty students (Sass et al., 2012) are associated with increased productivity gains over time. These findings have several implications for low-performing schools. Perhaps most importantly, productivity increases for teachers are influenced by the learning communities in which they work and not merely by their individual development, another reason to attend to working conditions in low-performing schools.

Incentives and Teaching Development

A well-developed theory suggests that performance incentives may induce behavioral responses that increase productivity (Lazear, 1995). This theory has been applied in education in the form of performance pay in an attempt to alter teacher behavior to improve student achievement. Typically teachers are provided with a one-time bonus for reaching achievement gain benchmarks. Research on these pilots usually finds they are ineffective in increasing student achievement. For example, recent experiments in Nashville, New York, and Chicago indicate that performance incentives in these cities were ineffective in increasing student test outcomes (Springer et al., 2010; Fryer, 2013; Glazerman & Seifullah, 2012). These experiments share some design elements that may explain the null effect of incentives on achievement. They are each designed to be short-lived experiments with incentives in the form of one-time bonus payments and teachers are rewarded solely on student achievement gains with little or no support to help them. Thus, teachers may not know how to respond to realize the bonus, or have sufficient incentive to do so for a short-lived bonus opportunity.

In contrast, IMPACT, the District of Columbia teacher evaluation system referred to earlier, introduced strong teacher dismissal and financial incentives that were linked to an index of multiple measures of teacher performance. IMPACT is an at-scale system, refined over several years, that provides teachers with supports to realize their goals. Dee and Wyckoff (2013) find that IMPACT led teachers who were subject to the strongest incentives to meaningfully improve their performance. Minimally effective teachers—those who would be dismissed if their performance did not improve to "effective" in the following year—improved by 27 percent of a teacher-level standard deviation. Similarly, teachers who had been rated "highly effective" and who would receive an increase to base-pay of about $25,000 if they were rated highly effective in the following year, improved by 24 percent of a teacher-level standard deviation. The experience of IMPACT suggests that substantial incentives in combination with supports embedded in a system sustained over years may improve teacher effectiveness.

Lastly, schools may also be able to improve the overall productivity of teachers by strategically assigning teachers. Several recent studies show that school principals move less effective teachers into non-tested grades and subjects (Chingos & West, 2011; Cohen-Vogel 2011; Fuller & Ladd, 2013; Grissom, Loeb, & Kalogrides,
2013) and distribute students to less experienced teachers in tested grades and subjects more equitably (Grissom et al., 2013). These strategic personnel moves are designed to increase measured school performance but as Grissom et al. point out, their effect on overall performance in both tested and untested grades and subjects remains unclear and depends upon persistence and timing of effective teaching and the match quality of the newly assigned roles.

Summary

Low-performing schools, which disproportionately house the least effective teachers, have been disadvantaged by a system that, until recently, paid little attention to ways of improving teacher effectiveness. Research is beginning to explore the specific skills that lead to effective teaching and how they can be best conveyed to teacher-candidates and teachers. Clearly, average teacher effectiveness increases markedly and steadily during the early years on the job. A relevant policy issue is what is it that teachers learn during these early years and could this learning occur during the preparation period to the benefit of teachers and students?

Conclusion

There is no more important school or district level educational input than effective teaching. In this chapter we have reviewed features of the teacher labor market that work against the equitable distribution of effective teachers and discussed how policies targeting compositional changes to the workforce and systematic efforts at teacher improvement that could benefit students in low-performing schools.

However, despite considerable research and policy attention devoted to attracting, selecting, retaining and improving teachers to develop an effective workforce in schools with low-performing students, much remains to be learned. Here are a few of what we consider to be the highest-priority research questions:

- What practices are most effective in improving teacher effectiveness?
  - What structure, content, and timing of teacher preparation is most effective in producing teachers who will be effective in classrooms with low-performing students, and does this differ across teaching contexts?
  - What content and combination of individual and group professional development are most effective?
  - Can teacher evaluation systems be structured to meaningfully improve teachers and teaching? If so, what are the mechanisms?

- To what extent is teaching effectiveness unevenly distributed across schools as suggested by value-added research?
  - Do we have good statistical identification across teaching contexts?
  - Do other measures of effectiveness produce similar disparities?

- How can low-achieving schools selectively attract and retain the most effective teachers?
  - What mechanisms realize this outcome most efficiently—compensating salary differentials or altered working conditions?
  - How can working conditions be altered to be attractive to high-performing teachers? For example what is the potential for teams of proven teachers and a proven principal to “take over” a school and transform the learning culture?
  - What is the effect of more rigorous tenure standards on teacher effectiveness and student achievement? Here it would be interesting to engage in planned variation across several
We conclude with an observation about research and policy. Many states and school districts implement policy before there is strong causal research. In the press for addressing chronic low achievement, policy preceding evidence is understandable. However, it is problematic when promising policies are abandoned before rigorous research can appropriately assess their effects. Researchers and policymakers should partner in the design and implementation of policies to ensure conditions for rigorous evaluation. Additionally, policies should be left in place until the hypothesized behavioral responses have time to be realized.

Notes

1. Attempts to employ value-added to compare teacher effectiveness across schools and districts confront some methodological challenges. Value-added estimates of teacher effectiveness must distinguish teacher effectiveness from the differential ability of students to improve. To control for differences in student ability to achieve, value-added estimates frequently control for student attributes, such as poverty. However, if, as suggested above, teachers sort themselves by ability in response to student poverty, controlling for poverty may control away some of the differences in teacher effectiveness.

2. There is some evidence that this is not the case (Ballou, 1996).

3. See Dee and Wyckoff for a summary of IMPACT.

4. Only 11 states have a probationary period longer than three years (National Council on Teacher Quality, 2012).

5. Koedel, Parsons, Podgursky & Ehlert (2012) caution that unless standard errors are appropriately clustered, they will be underestimated and the significance of program effect differences exaggerated.

References


Cambridge, MA: Harvard University.
Effects of Educational Policies on Disadvantaged Students

From Desegregation to Accountability

DOUGLAS N. HARRIS

Introduction

Improving educational outcomes for racial minorities and students from economically disadvantaged households has long been a central objective of federal and state educational policies. This chapter examines the effects of the four broad government strategies that have been implicitly or explicitly targeted to the educational outcomes of disadvantaged students: desegregation, funding equity, and test-and market-based accountability.

The federal government’s involvement in educational equity arguably started with the U.S. Supreme Court’s 1954 *Brown v. Board of Education* decision to end racial segregation in schools. Although federal policymakers were initially reluctant to support the decision, they eventually helped to implement and build on it with complementary policies such as the Elementary and Secondary Education Act (ESEA) of 1965. Part of the rationale behind the *Brown v. Board* Court decision was that the segregation of schools leads to resource inequity. While desegregation and the initial incarnation of ESEA helped to increase resources for minorities and low-income groups, the vast majority of funding still came from local property taxes. With housing still heavily segregated within and across districts, this meant that school spending also remained highly unequal. Two decades of state-level lawsuits and legislation over funding “adequacy” followed, further reducing spending gaps in many states (Card & Payne, 2002).

The most recent incarnation of ESEA, commonly known as No Child Left Behind (NCLB), takes a different approach, reorienting federal education policy away from access and resources toward accountability for educational outcomes. NCLB requires that students of various racial, income, and programmatic subgroups make “adequate yearly progress” toward state-determined proficiency standards. Thus, the law is specifically designed to focus attention on students who have historically achieved lower educational outcomes. While almost all states are now operating under federal waivers from NCLB, and significant changes in the law seem likely on the horizon, the focus on the academic proficiency of disadvantaged subgroups appears likely to remain at the law’s core. Underlying NCLB and other forms of test-based accountability is the view that schools have insufficient incentives to use resources wisely.

A fourth popular policy stream, market-based accountability, also increases the incentive for schools to use resources more efficiently by inducing schools to respond to parent demands. Charter schools and vouchers remain two of the most popular forms of school choice and disproportionately serve low-income and minority students (Zimmer et al., 2009).

While less explicitly focused on helping the disadvantaged, test-and market-based accountability still share with desegregation and funding reforms an implicit and disproportionate focus on the educational experiences of minority and low-income students. Because children with low test scores are disproportionately disadvantaged, any pressures created by test-based accountability to increase test score levels are likely to be strongest for these students. Similarly, depending on the details of the program, children from low-income families could potentially benefit the most from market-based reforms because the traditional public education system offers them few viable schooling options.
This chapter examines evidence about the effects of these four broad policy strategies on disadvantaged students in the United States over the past half century. Much of the extant research reviewed here on desegregation, resources, test-based accountability, and market-based accountability focuses on how policies affect student achievement as measured by scores on standardized tests. This focus on achievement is understandable both because of the primacy given to student achievement in current reform efforts and because it is one predictor of long-term success (Currie & Thomas, 1999; Neal & Johnson, 1996). This chapter also refers to the relatively small number of rigorous studies that measure the long-term effects directly in the form of graduation rates, college-going, employment, and wages.

While other chapters in this volume address some of the same policies and outcomes, they typically focus on average effects and efficiency, while this one focuses on equity.

**Desegregation**

Evidence related to desegregation comes from a number of sources, including early experiments to desegregate schools, housing experiments designed to alter neighborhoods, and studies of peer effects. The research generates findings related to both short-and long-run consequences.

**Desegregation Experiments and Black Achievement**

During the early period of the school desegregation movement, especially the late 1960s and 1970s, courts required schools to desegregate, usually by busing K–12 black students from their homes to schools serving predominantly white students. In some cases, children were reassigned to more integrated schools by lottery, thus providing the basis for experimental evidence on the effects of desegregation.

One of the most studied programs is the Boston METCO program, the largest and longest-running metropolitan desegregation effort in the nation that has been operating continuously since 1966. More black students applied to switch schools than there were slots available in the program’s early years, so black applicants were randomly assigned to be bussed to whiter schools in the suburbs (the treatment group) or to remain in a school in Boston (the control group). Both the treatment and control groups were given academic achievement tests before and after busing took place.

Many of the students in the control group were siblings of those who were bused, but who had been denied the right to transfer. Comparing sibling pairs of treatment and control students, Armor (1972) finds no gain in student test scores for the students bused to the suburban schools. Subsequent analyses of the METCO program, based on different methods and considering students over several different decades, have confirmed Armor’s finding (Angrist & Lang, 2004).

Studies of similar programs come to different conclusions. In 1984, the National Institute of Education (NIE) commissioned seven extensive reviews of the topic, each considering essentially the same group of studies (Armor, 1984; Cook, 1984; Crain, 1984; Miller & Carlson, 1984; Stephan, 1984; Walberg, 1984; Wortman, 1984). An independent review by Crain and Mahard (1983) found that the four studies of desegregation that were based on random assignment of students to the treatment and control groups were more likely to find a positive and significant effect and that the (weighted) average effect size was large by traditional research standards. Of the 21 reported effects from this small group of studies, 17 effects were positive and significant. Bused students made large average gains of 0.235 standard deviations, or 8 percentile points, per year compared with the control group. Other longitudinal studies with good control groups also yielded positive, but smaller, effects of 0.084 standard deviations. Although researchers differ in their interpretations of these studies, depending in part on which research method they favor, all of the NIE-commissioned reviews concluded that these desegregation programs raised the test scores of the participating black students (Schofield, 1995).

Several patterns emerge in these studies concerning the design and implementation of desegregation programs. School district-initiated or “voluntary” desegregation is more likely to generate positive effects than desegregation resulting from court orders (Bradley & Bradley, 1977; Cook, 1984; Schofield, 1995; Stephan, 1984). Indeed, the review by Bradley and Bradley (1977) suggests that all of the voluntary desegregation programs
raise black achievement. Effects may also vary depending on whether families volunteer to participate, though there appears to be no evidence on this topic.

Students who are moved to a desegregated school during their elementary school years also seem to benefit more than those desegregated later in their school careers (Cook, 1984; Crain, 1984; Crain & Mahard, 1983; St. John, 1975; Stephan, 1984). Harris (2006) suggests two possible reasons for this pattern. First, elementary students may experience fewer social adjustments than older students who have spent more time in segregated settings. Also, secondary schools are more likely to track students by academic ability (and perhaps race), which could reduce the benefits of desegregation for minorities.

**Housing Opportunity Experiments**

Two major programs during the past two decades, Gautreaux and Moving to Opportunity (MTO), have attempted to move low-income families out of racially and economically isolated housing projects and neighborhoods to more racially integrated or higher-income neighborhoods. Given the correlation between neighborhood characteristics and common measures of school quality, these programs may also have moved students into better schools.

In Gautreaux, a nonprofit organization relocated more than 7,000 African American families living in Chicago public housing to other parts of the metropolitan area, some within the city and others to suburbs. Because the families were not randomly selected for relocation, studies of this program are subject to concerns about selection bias. Nevertheless, the available data indicate that those who moved to the suburbs were much less likely to drop out of school and had higher rates of college attendance compared with students in the inner-city schools from which the students moved (Kaufman & Rosenbaum, 1992).

The apparently positive outcomes of the Gautreaux program inspired the U.S. Department of Education to initiate the more ambitious MTO experiment. Under this program more than 9,227 low-income and mostly minority families in Baltimore, Boston, Chicago, Los Angeles, and New York were randomly assigned to receive one of two “treatments.” One treatment group was given only Section 8 federal housing vouchers to be used wherever they could find suitable private-sector housing. The second treatment group was given the same vouchers combined with counseling and a restriction that the vouchers be used only in neighborhoods where the poverty rate was lower than 10 percent.

Ludwig, Ladd, and Duncan (2001) examine MTO effects for the Baltimore participants using data from two school district-administered achievement tests. Their results suggest statistically significant differences between the experimental and control groups for students who entered the program when they were younger than 12 years old. Compared with the control group, these treatment-group students were 17.8 percent more likely to pass the state standardized test and scored 7 percentile points higher (equivalent to 0.17 standard deviations) on a district test. Consistent with the effects on desegregation, no effects on state test scores emerged for those students entering the program after age 12.

A second study of the MTO includes all five cities and used different outcome measures. Instead of district achievement tests, the results reported in Sanbonmatsu et al. (2006) are based on a test that was not part of the cities’ standards and accountability systems. In contrast to Ludwig, Ladd, and Duncan (2001), these authors find no achievement effects for students in any age group.

Two possible explanations could account for the different MTO results. First, because the test used by Sanbonmatsu et al. (2006) was not aligned with state standards or accountability, it may not have measured whether students were learning the curriculum taught in schools. Second, the data in Sanbonmatsu et al. (2006) show that the movement of families under the MTO resulted in only small changes in the characteristics of the students’ classmates. The students’ classmates who were racial minorities or who were eligible for free or reduced price lunches declined by only 3 to 6 percentage points and the mean percentile test score rank increased by only four points, changes that may be insufficient to generate significant differences in student learning. In any event, the multifaceted nature of housing moves means that researchers are unable to distinguish neighborhood and school effects.

**Peer Effects and Achievement**
Central to arguments for racial or economic desegregation is how students’ peers affect their educational outcomes. The classic paper on this topic, Jencks and Mayer (1990), provides two main models for why having advantaged peers may improve a disadvantaged student’s educational outcomes. With the “epidemic” or contagion model, children’s behavior, such as the propensity to do homework, has positive spillovers on other students in the classroom, so children in classrooms with motivated and high-achieving peers are likely to perform at higher levels. In contrast, the “institutional” theory, part of the core of the Brown v. Board decision, highlights the fact that districts serving advantaged students are likely to have more resources than other schools. Within districts, similar inequalities often emerge across schools because of the greater political power of the higher-income parents.

A different possibility, labeled by Jencks and Mayer’s (1990) the “relative deprivation” theory, posits that some children may become more frustrated and put forth less effort in the presence of advantaged peers because of their relative social position. A variation on the relative deprivation model, “cultural conflict,” posits that students who are low in the perceived social hierarchy choose to disassociate from the dominant group and form alternative peer groups that reject the dominant group’s preferences and behaviors. These and other related theories (see Harris (2010) for a comprehensive review) suggest that being in a school or classroom with advantaged peers could be detrimental. As we show below, the evidence suggests that any harmful effects of having advantaged peers are outweighed by the benefits.

A growing number of studies use longitudinal analysis of student test scores—sometimes called “value-added” modeling—to address the inherent challenges in identifying causal peer effects. The value-added studies also provide more comprehensive information than do the previously discussed school and housing integration studies. First, these studies make it possible to disentangle the separate roles of peer race, peer income, and peer achievement. Second, the newer evidence fills an important gap in the studies of school desegregation by considering not only the effects of peers on minority students, but also the effects on white students. While the primary purpose of desegregation is to address the needs of minority and disadvantaged students, it is important to know whether benefits for targeted students come at the expense of other students.

A final benefit of value-added studies of peer effects is that they avoid the possible confounding influence of disruption that occurred during the early era of desegregation that may have altered the measured effects of the programs (Hanushek, Kain, & Rivkin, 2002). Being one of a handful of black students bused to a suburban school—in some cases accompanied by newspaper headlines, picket lines, and National Guard troops—likely influenced the educational environment in detrimental ways. These disruptions might be avoided with other forms of integration, or might be absent in the present era where such controversy has diminished. The following discussion summarizes findings from five major value-added studies of peer effects.

The first consistent finding is that having advantaged peers results in better outcomes for minority students. This result emerges from each of the three studies that estimate peer effects by race: one study conducted in North Carolina (Cooley, 2006) and two studies in Texas (Hanushek, Kain, & Rivkin, 2002; Hoxby, 2000). These studies also find that the benefits of advantaged peers for whites are smaller than the benefits for minorities. A fourth study, by Burke and Sass (2006), finds that advantaged peers are beneficial, but this study does not estimate the effects by racial subgroup.

Because minorities benefit from having advantaged peers, and more so than do white students, these studies suggest that desegregation could improve outcomes for nonwhite students and increase average achievement simultaneously. However, some nuances or complexities in these findings are worth noting. In particular, members of each racial group appear to be influenced most by peers of their own race. Hoxby (2000) finds that peers are influenced more by the percentage of students in their specific racial group (e.g., blacks are more negatively influenced by an increase in percent of black students than by an increase in the percent of Hispanic students) and by the achievement of same-race peers. Hanushek et al. (2002) find negative effects from peer percent black and Hispanic that are larger within-race, as does Cooley (2006), though she combines blacks and Hispanics into a single “minority” category. This general finding of larger within-race effects may be explained by the fact that all students more closely identify with, and are therefore more affected by, students who are similar to themselves, what Harris (2010) calls “group-based contagion.”

It is also possible that peer influences depend on the individual student’s initial test score. Burke and Sass (2006) find that achievement peer effects occur for students throughout the distribution of test scores, and the effects are largest for students who are among the lowest 20 percent of the test score distribution, and smaller
for high-scoring students. Cooley (2006) finds this same result for white students, but that the reverse is true for minorities—high-scoring minorities benefit the most from higher-scoring peers. While Burke and Sass (2006) do not break out their results by racial subgroup, it is likely that their results are driven by the white students who represent the majority of their sample, making their results consistent with those of Cooley.

Inconsistencies also arise in the specific ways that peers matter. First, Hanushek et al. (2002) find that peer influences are most closely related to peer percent minority and are unrelated to peer achievement. In contrast, Cooley (2006) and Hoxby (2000) find that both peer percent minority and peer achievement are associated with individual achievement. Second, Vigdor and Nechyba (2007) find that having more minority peers actually has a positive impact within racial subgroups (e.g., blacks benefit from having more black peers), after controlling for peer achievement. Harris (2006) finds this same result in a larger sample of students, though he has access to only subgroup-by-race data and for a limited number of years. These inconsistencies in findings may reflect differences in the states being studied, as well as differences in methodology.

For at least two reasons, these estimated peer effects may not be fully causal in nature. Vigdor and Nechyba (2007) find that the apparently positive effects of advantaged peers are eliminated when unobserved differences in teachers are taken into account. This contrary finding may not be as significant as it seems, however, because part of the benefit of having advantaged peers may be that teachers are generally more willing to teach advantaged students. The Vigdor and Nechyba results do not allow for these “institutional” benefits. None of the other studies account for differences in teachers in this way.

Some of the estimated peer effects are also implausibly large. For example, Hoxby’s (2000) estimates of peer effects imply that moving a black student from a 100 percent black school to a (slightly less than) 100 percent white school would raise individual achievement by an additional 0.7 standard deviations in a single year. Thus her results suggest that the full black-white achievement gap of 0.7 to 0.9 standard deviations could be eliminated in a single year simply by moving students to different schools. There is considerable evidence of substantial achievement gaps even in schools that are mainly white, as highlighted by the well-known Shaker Heights example (Ogbu, 2003). If blacks really could catch up in majority-white settings, then these within-school gaps should be considerably smaller than they are. Nevertheless, given the consistency with which researchers find benefits of more advantaged peers, including through experiments, it seems likely that at least a portion of these effects are causal.

**Long-Term Effects of Desegregation on Blacks**

These effects of racial integration and peer influences on achievement are important because achievement is a predictor of long-term success. Integration might also affect long-term outcomes in ways unrelated to achievement. As described by Wells and Crain (1994), integrated schools or neighborhoods may give black students the confidence and ability to cope with situations involving whites, who make up the majority of people in college and the professional workplace. Integration also establishes social networks that provide both information and personal connections that help students as they navigate their educational and career paths.

Again, the most convincing studies are those that address selection bias through random assignment and track the randomly assigned students over long periods of time. In the Boston METCO example, Armor (1972) compared the treatment and control groups up to two years after high school graduation. Even though there were no achievement gains, Armor finds longer-term educational gains: 84 percent of the bused students went on to college after high school compared with 56 percent in the control group. More modest evidence of long-term effects is found in the MTO housing program. Ludwig et al. (2001) find some reduction in dropout rates, but their estimates are significant in only one specification, and only then at the 10 percent level of significance.

Unfortunately, very few of the desegregation experiments involved long-term follow-up of students. It is therefore worth considering other types of analyses, including correlational analysis. Using the High School and Beyond data, Rivkin (2000) finds that having more white classmates is slightly negatively associated with black test scores, but he and Grogger (1996) both find a positive association with black educational attainment and wages, other things being equal. The effects on long-term student outcomes are lower, and in some cases negative, in urban school districts that were under court-ordered desegregation.

Compared with the earlier experiments, it is somewhat more difficult to draw clear conclusions from the
studies of longer-term effects because there are fewer of them and many rely on correlational analysis. Nevertheless, gains for black students, as measured in the most convincing studies of integration, are almost universally positive.

Despite the limitations of individual studies, they collectively paint a consistent picture: Having more advantaged peers is beneficial, especially for minority students and at the elementary school level where desegregation was voluntary. Peer influences are complex, however. Desegregation is not always beneficial, individuals are affected most by peers of the same race, and the effects vary depending on whether the individual student is low- or high-achieving.

Having more minority peers may generate inconsistent effects on student achievement because what sometimes appears to be a race effect may actually reflect the socioeconomic status (SES) of the peers. While no studies provide convincing results regarding this hypothesis, the role of peer family income is arguably more important for policy because recent U.S. Supreme Court decisions have limited the ability of school districts to assign students to schools on the basis of race. As a result, at least 40 school districts have already changed strategies and now assign students to schools on the basis of income (Kahlenberg, 2006). Reardon, Yun, and Kerlaender (2006) show that income-based desegregation may not generate much racial integration because: (a) whites are a much larger share of the population and many of them also have low incomes; and (b) individual districts are racially homogeneous, so there are often few opportunities for within-district racial integration. Income-based desegregation may still be beneficial for disadvantaged students given the above evidence on peer effects, but would not apparently address racial imbalances.

Resources and Funding

One of the main ways in which blacks benefited from desegregation was through increased school resources. The federal Title I program was instituted in part to encourage states to follow Brown v. Board of Education. Even before this groundbreaking decision, states began increasing resources in predominantly black schools to head off lawsuits and address growing political power of the black community during the Civil Rights era.

The changes in resources were dramatic: Card and Krueger (1992) show that resources for black students increased substantially during the post–World War II era in Southern states. In 1946, there was a 42 percent gap in salary paid to teachers in predominantly black versus white schools, but this gap was nearly eliminated by 1965. Likewise, the average pupil-teacher ratio in black schools was 22 percent greater than that in white schools in 1946. This gap decreased to 8 percent in 1965, which is noteworthy considering the absolute reductions in pupil-teacher ratio for white students during this same time period. Boozer et al. (1992) show that these trends continued and that the pupil-teacher ratio was essentially identical across racial/ethnic groups by 1989.

More recent evidence focuses on the distribution of funding across districts based on student socioeconomic status. Baker, Sciarra, and Farrie (2012) find that spending levels are now essentially the same in districts with high percentages of students in poverty versus those with no poverty. These figures mask funding inequities at the school and classroom level, however. More advantaged families within lower-income districts use their political power to direct more resources to particular schools, for example, and it is more difficult to attract teachers to low-income schools (Hanushek, Kain, & Rivkin, 2001). Within schools, advantaged students are also “tracked” into the more rigorous courses that facilitate success in college.

A well-known debate has occurred about whether funding generally influences student achievement (see Rice & Schwartz, this volume). If funding is important in general, it is likely to be more important to disadvantaged students because of the concept of diminishing returns. Family and school inputs together represent the total inputs devoted to increasing outcomes, and since disadvantaged students have lower family inputs by definition, they should benefit more from school inputs because they are on the steeper part of the production function (Harris, 2007).

The original federal effort to increase resources devoted to low-income students, Title I, has limited supporting evidence, however. Syntheses of early work suggested that Title I may have had a modest positive impact on student test scores (Borman & D’Agostino 1996), but these early studies rely on correlational analyses. More recent studies have used more rigorous methods, albeit only in individual schools and districts,
and find no or even negative effects (van der Klauuw, 2008; Matsudaira et al., 2012).

Other studies have looked at funding shifts beyond Title I. Grissmer et al. (2000) conclude that while extra resources have probably had very small positive effect on average, “additional resources have been most effectively used for minority and disadvantaged students” (p. 41). Card and Payne’s (2002) national state-level analysis suggests that efforts to increase funding equity reduced the gap in SAT scores between students whose parents had low levels of education relative to those with higher levels of education. However, this study has numerous limitations and even the authors describe the results as “tentative” (p. 80).

The uncertainty over the effects of funding is likely to remain because it is difficult to clearly separate correlation from causation. Even when natural experiments lead to sharp changes in funding, the effects are often delayed and/or offset by reductions in other funding. Funding may also be complementary with other policy changes, such as those discussed in the following sections, that are designed to induce schools to use existing resources more efficiently.

**Test-Based Accountability**

The national perception in the late 1970s was that desegregation, funding equity, and other changes in school policy had failed or even backfired. Average college entrance exam scores appeared to be on the decline, and the general public believed that schools had lowered academic standards, making schools more concerned with keeping students from dropping out than with making sure they could read and write (Harris & Herrington, 2006).

*A Nation at Risk* (NAR) gave voice to these student achievement concerns (National Commission on Excellence in Education, 1983), and state policymakers made a concerted effort to raise academic standards. Increased accountability based on standardized tests aligned to these standards soon followed. Today, test-based accountability takes two general forms. In its early incarnations, testing was used to hold students accountable through promotion and graduation exams while, more recently, it has evolved to hold teachers and leaders accountable for student results with school report cards and associated sanctions and rewards.

**Evidence on Promotion/Graduation Exams**

One response to the perceived reduction in academic standards was to end “social promotion”—to ensure that students who earn high school diploma have met basic academic requirements. In at least 25 states, these calls for reform have been answered with promotion and graduation exams (PGEs) that require students to achieve a minimum score on a standardized test before being promoted to the next grade, or in the case of high school seniors, before being allowed to receive a diploma (Center on Education Policy, 2011). Students who do not pass the graduation exam, initially or in subsequent attempts, cannot graduate from high school. They generally either drop out of school or settle for less valuable Graduate Equivalent Degrees (GEDs).

PGEs are more common in states serving more disadvantaged students. While 69 percent of the nation’s students are subject to graduation exams, the figures are higher for blacks (71 percent), Hispanics (85 percent), low-income students (71 percent), and English Language Learners (83 percent) (Center on Education Policy, 2011). Disadvantaged students who take the test are also more likely to be affected by the policy because their lower average scores make them more likely to fail the test. They may benefit from the additional time and resources associated with being held back (the resource effect) or from being induced to work harder to pass the test (the incentive effect), as has been found in two studies of New York’s Regents Exam (Bishop, 1998; Alexander, 2000). Alternatively, the disappointment and/or loss of friends and social networks might make students worse off. Promotion exams in earlier grades give students time to catch up and pass the test. Graduation exams allow less time to catch up, so high school students may ultimately never pass the test and may lose the chance to receive a high school diploma.

In Chicago, the district’s policy required that students scoring below the required level in grades three, six, and eight take summer school, retake the exam, and earn a passing score before being allowed to advance to the next grade. Roderick et al. (2002) found that promotion exams generally raised achievement in grades six
and eight, but not grade three. Further, while schools with large portions of high-risk students appear to benefit the most in terms of reading in all three grades and in third-grade math, the low-risk students benefit the most in sixth-and eighth-grade math. Miller et al. (2002) report that these gains translated into better high school outcomes. Nagoaka and Roderick (2004) find no effect of promotion exams on the achievement of those students who failed and were held back, suggesting that gains are most likely due to the incentive to increase performance rather than the resource effect of the extra time in school for those students held back.

Two studies of the Chicago reforms examine more directly the equally important issue of why accountability might raise test scores. An in-depth, small-scale study by Roderick and Engel (2001) based on interviews with students found that, as a result of accountability, two-thirds of elementary students paid greater attention and worked harder on their school work and experienced greater pressure and support from teachers. The other one-third of students, however, was further behind academically and therefore faced greater difficulty in passing the standardized test. In sum, students far below the cutoffs did not respond with additional effort.

Jacob (2005) confirms that accountability led to higher test scores in Chicago, especially for middle-and, to a lesser degree, upper-elementary students. Like Roderick and Engel (2001), he also finds that this gain for middle-and upper-elementary students was accomplished in part through increased student effort. Additionally, Jacob identifies other strategic behaviors that contributed to the higher scores: Schools increased the number of students placed in special education and shifted the curriculum away from nontested subjects such as science and social studies. In short, these findings suggest that accountability produces both substantive improvement and strategic responses of questionable educational value.

Identifying appropriate comparison groups is always a challenge for causal inference, but the problem is distinctive when studying the effects of PGEs on achievement. The basic issue is that retention involves separating students from their grade cohorts so that they are no longer taking the same standardized tests in the future. Schwerdt and West (2013) describe the two main alternatives for addressing this problem and show how the assumptions differ when comparing retained students to their promoted peers versus their new cohort. It appears that the differences in assumptions are important because the two methods have yielded different results even with the same data, sample, and general methodology (regression discontinuity). Winters and Greene (2012) study the effects of Florida promotion exams using one approach and find positive effects, while Schwerdt and West (2013) use the other method and find no effects. While both approaches require assumptions, those in the Schwerdt and West paper seem more realistic.

Overall, the evidence tends to support some achievement gains from PGEs, though the contrary results from Schwerdt and West (2013) and the hard-to-explain pattern of effects found by Roderick et al. (2002) raise noteworthy concerns.

Promotion/Graduation Exams and Long-Term Outcomes

Regardless of the short-term effects on achievement, it is possible that PGEs improve long-term outcomes by providing students with human capital that improves life outcomes. Again, this benefit might arise even for students who fail the exam and/or those who drop out, as long as the incentives of the PGEs induce them to work harder in school. The existence of PGEs may also improve the quality of the “signal” that high school diplomas send to employers about student knowledge and skills, thereby improving the Efficiency of the labor market for all workers (Bishop & Mane, 2001).

Almost all empirical studies on the subject find that PGEs increase the probability of dropouts, especially for students who have low academic achievement. Three studies have used the nationally representative data and they reach this conclusion (Dee & Jacob, 2006; Baker & Lang, 2013; Hemelt & Marcotte, 2013). The negative effects are worse for African Americans. As Hemelt and Marcotte (2013) put it: “States that adopt exit exams with no alternate pathway can expect to see increases in the dropout rate of African-American 12th graders of more than 50 percent for males and 65 percent for females” (p. 339). The increase in dropouts is especially pronounced in states that lack an alternative pathway to graduation. Other studies in Chicago and Texas also find negative effects on graduation (Allensworth, 2004; Martorell, 2005).

So far, the news on PGEs is mixed, suggesting that possible achievement gains are offset by reduced rates of graduation. The picture becomes more confused when we consider college enrollment and labor market outcomes. Martorell (2005) finds that PGEs reduce the probability that students just below the cutoff will go on
to higher education. This is what we would expect, given that such students are less likely to obtain the high school degrees that are often necessary for admission to college. However, Dee and Jacob (2006) find no effect on college attendance.

The effects of PGEs on employment outcomes are also inconsistent and difficult to explain. Dee and Jacob (2006) and Baker and Lang (2013) find that while PGEs have no employment or earnings effect, while Bishop and Mane (2001) do find a positive effect on earnings. Perhaps the only consistent finding is that PGEs have positive effects on Hispanic females, in terms of both college attendance (Bishop & Mane, 2001; Dee & Jacob, 2006) and employment (Dee & Jacob, 2006). While effects on subgroups are more likely to be spurious in general, the consistency of this finding across studies and analyses is potentially important.

Taken together, this evidence suggests that PGEs may have some positive effects in terms of achievement, but the clearer effect is to redistribute educational credentials to those who pass the tests from those disproportionately disadvantaged students who fail them. The mix of college and employment effects is perplexing and probably, despite the best efforts of the researchers, do not reflect causal relationships. With improvements in the ability to track students over long periods, future research will likely better inform our understanding of the long-term benefits and costs.

**Test-Based Accountability for Schools**

PGEs place the onus on students: if they fail the exams, they are not promoted or do not graduate. Other more recent forms of federal accountability place the burden on the schools themselves and school personnel. If schools do not reach specific achievement thresholds, schools are given well-publicized negative labels, forced to accept closer scrutiny, and occasionally required to replace the principal or reorganize the school.

The name of the federal law, No Child Left Behind (NCLB), implies a direct focus on the disadvantaged, as does the law’s requirement that schools report results by racial and income subgroups and show improvement for these groups. More indirectly, but perhaps no less significantly, NCLB set a goal of proficiency (albeit one defined at the state level) for all students. While the goal of 100 percent proficiency is no longer the focus, accountability systems still attend to this bar—a bar that disadvantaged students are much less likely to reach.

A growing body of research provides mixed evidence of success. In their study of pre-NCLB state test-based accountability, Carnoy and Loeb (2002) find that an index of state accountability, measuring the overall level of external pressures on school personnel from a variety of policy mechanisms, is often positively associated with state student achievement.

Moreover, these gains are largest for blacks and Hispanics. Building on these results, Harris and Herrington (2004) separate the same accountability index into its component parts (PGEs, school report cards, and associated report card incentives) and additional state policy characteristics (e.g., market-based accountability). Consistent with the frequently positive PGE effects discussed earlier, they find that PGEs most likely account for the positive effects found by Carnoy and Loeb (2002), Bishop, Mane, Bishop, and Moriarity (2001), and Raymond and Hanushek (2003) perform similar analyses for school report cards and find some evidence for the scarlet letter argument, though the analysis by Harris and Herrington suggests that these two studies may be capturing effects of PGEs and other policies that are excluded from their analyses.

Dee and Jacob (2011) study the effects of NCLB, leveraging the fact that, while implemented nationally, the variation in state accountability systems means that the federal law represented a larger change in accountability for certain states. The authors find that states whose accountability systems were most affected by NCLB saw either similar or in some cases larger National Assessment of Educational Progress (NAEP) achievement gains in math, but no effects in reading. The effects in math were noticeably larger for minorities and low-income students, though there were no clear subgroup patterns in reading. Using somewhat more recent data and slightly different methods, however, Lee and Reeves (2012) find less positive effects.

The NCLB direct incentives to increase achievement for racial minorities and low-income subgroups seem to have their intended effect of increasing achievement for these groups. Using data from North Carolina elementary and middle schools, Lauen and Gaddis (2012) find that the subgroup requirements induced schools who failed on the subgroup requirements to increase achievement specifically for these disadvantaged subgroups.

The students most affected by the proficiency focus are those near the proficiency thresholds, the so-called
bubble students (Neal & Schanzenbach, 2010; Rouse et al. 2013). Since states vary in how high they set their proficiency targets, it is difficult to generalize about how these rules affect disadvantaged students. Paradoxically, in states with low-proficiency bars, disadvantaged students might benefit from the bubble effect phenomenon for the simple reason that they have lower average scores and may receive more attention and resources, although some research suggests that low-scoring students benefit even when growth measures are used instead of proficiency to hold schools accountable (Ladd & Lauen, 2010).

Schools also try to reclassify low-performing students into categories that are not counted in accountability metrics (Figlio & Getzler, 2006; Jacob, 2005; Cullen & Reback, 2006) or by subjecting them to longer disciplinary suspensions near testing dates (Figlio, 2006). These moves to exclude students from performance calculations are likely to disproportionately apply to disadvantaged students, although accountability laws are increasingly aimed at preventing such exclusions.

What can we conclude from this evidence on high-stakes testing? First, it seems clear that stringent accountability increases average math scores and that the promotion and graduation exams contribute to this result. But the picture becomes less clear when we focus on disadvantaged students and on the full range of outcomes. While both student- and school-focused accountability do seem to increase test scores for minorities, these benefits are apparently offset by losses in high school graduation. In this sense, the benefits depend not only on which student groups researchers focus on, but also on which outcomes. For a more complete discussion of test-based accountability, see Figlio and Ladd (this volume).

**Market-Based Accountability**

The early 1990s saw not only the expansion of test-based accountability but also various experiments to increase parental choice of school and market accountability. This section focuses on two specific policies that increased choice, especially for disadvantaged students: charter schools and private school voucher programs. Although both are funded from government revenue, the two differ in how the money is used. Charter schools, which are chartered by a public agency but managed by private organizations, are subject to fewer regulations than are traditional public schools. Public funds for school vouchers allow students to attend private schools with limited government oversight. Compared with charter schools, vouchers are available to a much smaller number of students nationally, though they are targeted more directly at low-income and minority students. The oldest programs, in Milwaukee, Cleveland, and other cities, have been followed by much larger statewide versions implemented in states including Florida, and more recently Louisiana.

Charter schools serve a student population that is disproportionately minority and somewhat lower income than the nation’s school-age population as a whole (Center for Research on Education Outcomes, CREDO, 2013; Frankenberg, Siegel-Hawley, & Wang, 2010; Zimmer et al., 2009). Researchers have attempted to identify the effects of these schools using two main approaches. Angrist et al. (2011) leverage the fact that some schools are oversubscribed and must therefore randomly select students. Using this approach, they find that students end up with higher test scores if they are randomly selected to attend a charter school. Charter schools seem relatively more effective than traditional public schools in urban areas.

The lottery approach may be problematic, however, because the schools that are oversubscribed are likely to be of higher quality than the typical charter school—they are oversubscribed because they are good schools. The Center for Research on Education Outcomes (CREDO) uses a much larger, national sample of schools and compares matched pairs of students in charter and traditional public schools (CREDO, 2013). This second approach yields no difference in average achievement effects between charters and traditional public schools.

How vouchers affect the achievement of minority students has been a topic of considerable controversy. In a study of a lottery-based program in New York City, Howell and Peterson (2002) at first concluded that voucher effects were positive and statistically significant for blacks. However, a reanalysis of these data by Krueger and Zhu (2004) shows that this result is partially due to the researchers’ decision to omit a large percentage of students who did not have baseline data. The results were also sensitive to the way in which student race was coded. Nevertheless, a meta-analysis looking across a larger number of studies suggests that voucher effects are generally somewhat larger for black than for white students (Rinquist, 2013). Some people have suggested that vouchers could significantly reduce achievement gaps (Howell & Peterson, 2002), but even if we accept the
more optimistic estimates of voucher effect, they would still offset only a small fraction of the total achievement gap.

While at best charters and vouchers generate modest effects on achievement, evidence suggests they may generate some longer-term positive effects on college outcomes for some groups. In a follow up to the New York City analysis of vouchers, Chingos and Peterson (2012) find no average treatment effect, but this masks considerable variation across groups. They find that the null average effect combines a positive and substantively large effect on college outcomes for blacks, offset by a possibly much larger negative effect for white students. Positive effects on college outcomes are also found in a quasi-experimental study of charter schools (Zimmer et al., 2009).

Charter schools and vouchers programs might generate larger benefits for disadvantaged students if high quality schools were located closer to where the students live, if families could more easily afford to drive their children to school, and/or if families had better information about school quality (Bell, 2009; Harris, Larsen, & Zimmerman, 2014; Teske, Fitzpatrick, & Kaplan, 2007). These considerations highlight the importance of the design of choice systems, including the availability of busing and school quality reports.

An additional concern is that charters and vouchers might reverse the success of the desegregation era, at least in terms of racial isolation across schools (Frankenberg, Siegel-Hawley, & Wang, 2010). Choice programs could promote resegregation and racial isolation for a variety of reasons: families may prefer attending schools where students have similar racial makeup, certain minority groups may be particularly open to such alternatives because they are not well served in traditional public schools, or charter school operators may design and market their schools to attract particular groups (Glomm, Harris, & Lo, 2005). Given that charter schools and voucher programs often do not provide transportation, racial isolation could also result from the tendency of families to attend schools that are close to home.

Regardless of the reason, overall differences in the minority percentages of those who attend charter schools say little about the extent to which charters affect racial isolation. A valid estimate of how parental choices affect racial isolation requires student-level data so one can compare the demographics of sending and receiving schools. In a study of eight states where charter schools are common, researchers who used this approach found that, for students who moved from traditional public schools to charter schools, the racial compositions in the sending and receiving schools were similar on average, although with considerable variation across states and ethnic groups (Zimmer et al., 2009). This pattern suggests that the disproportionately large share of disadvantaged students in charter schools in many cases reflects the racial segregation of urban areas where those schools are located.

Given the evidence on race-based peer effects, an increase in racial isolation might well lead to lower student achievement. This outcome appears to have occurred in North Carolina where charter schools apparently have caused both an increase in racial isolation and in the black-white achievement gap (Bifulco & Ladd, 2007). Based on the other national evidence cited above, however, the North Carolina experience appears to be an outlier.

For additional discussion of market-based accountability, see Bifulco and Bulkley (this volume), and Zimmer and Bettinger (this volume).

**Conclusion**

The strategies used to improve outcomes for disadvantaged students have evolved considerably in recent decades—from providing more resources to minority students through desegregation and more direct school funding efforts, to accountability that creates incentives based on student outcomes for schools to use resources more wisely. This chapter shows that these various strategies have had different degrees of success.

Evidence from decades-past experiments in school integration, and more recent peer-effect studies, suggests that desegregation improves the achievement of disadvantaged students. Moreover, the evidence from peer-effect studies suggests that racial integration has little negative effect on the achievement of whites, thereby making it an attractive policy option. The increase in school resources devoted to disadvantaged students, much of which was driven by the desegregation and Civil Rights movements, also apparently benefited disadvantaged students.
These past successes do not mean that similar strategies would be equally successful today. Given the broad changes in the role of race in society, and the diminishing returns we expect from additional resources, it is reasonable to think that new strategies might be necessary to generate further reductions in achievement gaps. Growing fiscal pressures, underfunded pensions, the aging population, and associated health care costs limit the amount of likely additional funding for education and will require doing more without additional resources in the years ahead.

Is accountability now the best path forward for disadvantaged students? Unlike the effects of desegregation and resources, the effects of test-and market-based accountability have been more mixed. Test-based accountability seems to generate some positive effects on average achievement, but PGEs appear to generate some negative offsetting effects, and the negative effects are likely to be felt most by disadvantaged groups. Accountability that creates more-aggressive incentives for individual educators are newer and therefore more difficult to judge.

Market-based accountability is also fairly new and remains much less widespread than test-based accountability. Charter schools in particular are likely to see continued steady growth. Although the evidence suggests that a universal charter program would do little to reduce achievement gaps, it might help increase college outcomes. With respect to accountability reforms, we know little about how they will affect the outcomes of arguably greatest interest: high school graduation, college attendance, earnings, and other long-term outcomes. Although both charters schools and vouchers show some positive effects on college outcomes, future research will no doubt try to extend the analysis well into adulthood.

The patterns of policy effects described here are consistent with long-term trends in black-white achievement gaps (Harris & Herrington, 2006). The black-white achievement gap diminished dramatically during the 1970s and 1980s when desegregation and finance equity policies—both apparently effective—were most aggressively implemented. In the 1990s and 2000s, reductions in the black-white achievement gap stalled and income-based achievement gaps grew. The stagnation of the black-white gap overlaps the period during which the focus of policy shifted to test-based and market-based accountability, which have less research support. While the recent stagnation in achievement gaps might be a partial indictment of the accountability strategy, it is perhaps more accurate to say that desegregation and resource equity were the low-hanging fruit. Any effort to further narrow achievement gaps will no doubt pose a greater challenge.

Other strategies may also be pursued. New Orleans, Washington, DC, and other cities are pursuing much more intensive versions of test-based and market-based accountability in which all parents have an opportunity to choose among publicly funded schools, and schools are closed if they do not generate high enough test scores. Early childhood education is another option of growing interest (see Bassok and Loeb, this volume). Through experimentation and research, perhaps these and other strategies will help narrow educational inequities going forward.

Notes

1. Schwerdt and West (2013) show formally the assumption required for the Winters and Greene (2012) approach to be valid. However, the first assumption (“any effects of prior grade retentions do not fade out”) can be addressed by limiting the sample to those without prior grade retentions; also, the assumption that there is “no grade 3 specific age effect” can be addressed by adding flexible forms of age to the model. This leaves the assumption that there is no “grade 3 specific year-of-schooling effect.” This assumption is required, but a similar assumption is also required in Schwerdt and West’s own analysis; their method relies on the assumption of a vertical scaled test so that test score growth can be directly compared across grades. To address this, they impose the assumption that growThis the same across grades, i.e., students learn the same amount in every grade on average. Given these various assumptions, it is unclear whether the Florida program improves outcomes. See also Jacob (2001) for additional early analysis of PGEs.

2. The authors do not report this negative effect because of baseline imbalance within the control group, although only the tests for blacks and Hispanics are shown.

References


Section VII
Special Circumstances

SECTION EDITOR: DAVID H. MONK
Special Education

TOM PARRISH, JENIFER J. HARR-ROBINS, AND JAY G. CHAMBERS

Special education is an important, unique, and mostly growing component of public education in the United States that has become a major focus of attention in regard to public policy, appropriate service provision, and levels of spending. This chapter presents data on special education spending in the United States over time, describes the various ways in which special education services are funded, presents some of the primary policy issues associated with special education finance, and discusses some of the implications for possible future research and policy. Although the primary focus is on special education in the United States, many of the trends and issues presented in this chapter also pertain to Canada, Europe, and elsewhere around the globe (Naylor, 2001; McLaughlin & Rouse, 2000; OECD, 1995, 2005).

A Brief Historical Perspective

Although institutions for children with particular disabilities such as deafness and blindness were established in the early 19th century in several states, they operated almost entirely independently of the public school systems. Even after compulsory attendance laws were established later in the century, children with disabilities continued to be excluded from public schools. Although several states began to upgrade the educational programs and services publicly available to children with disabilities, such programs were still relatively few in number at the time of the civil rights movement during the 1960s. By 1966, congressional hearings revealed that only about one-third of the nation’s children with disabilities were receiving appropriate special education services (Verstegen, 1994). It was at this time that the parents of students with disabilities began to organize and demand educational services for their children (Mosher, Hastings, & Wagoner, 1979).

At the federal level, special education funding took root with the Elementary and Secondary Education Act of 1965 (ESEA). The ESEA and its subsequent amendments established grants for state-operated schools to serve students with disabilities and later included public schools. The watershed event for special education took place in 1975 when Congress passed the Education for All Handicapped Children Act (Public Law 94–142), subsequently renamed the Individuals with Disabilities Education Act (IDEA). This legislation laid out the blueprint for federal special education law as we now know it.

In addition to establishing funding for special education, the IDEA guarantees eligible students with disabilities the right to a free and appropriate public education (FAPE) and created due process procedures for ensuring FAPE. Each special education student is required to have an individualized education program (IEP) that specifies the special education program and services he or she is to receive as determined by a multidisciplinary panel of educators and related service providers. The law further states that these students will be educated with nondisabled children to the maximum extent appropriate to their needs, a concept known as the least restrictive environment (LRE).

Unique Aspects of Special Education

Students who qualify for special education differ from general education students in important respects, many of which have fiscal implications. First, students receiving special education represent a substantial—and
generally growing—proportion of public school enrollments. The latest federal data available for 2012–13 indicate that more than 6.5 million students, ages three to 21, have been identified for special education.\(^1\) The percentage of public school students identified for special education grew from 8.3 percent in 1976–77, the first year such data were collected, to 13 percent in 2010–11.\(^2\) While this growth was steady for nearly 30 years, reaching a peak of 13.8 percent in 2004–05, since then it has begun to decline somewhat. Along with expanding enrollments, special education services comprise a significant proportion—an estimated 14 percent in 1999–2000—of spending on elementary and secondary education (Chambers, Pérez, Harr, & Shkolnik, 2005).\(^3\)

Second, as noted above, special education students have a legal right under federal law to a free and appropriate public education (FAPE), an entitlement that distinguishes them from all other students. Unlike general education, the district is legally bound to provide services that are specified on a special education student’s IEP, and cost may not be a precluding factor. The IEP, which is a legal contract between the responsible school district, the special education child, and her or his family, creates an entitlement to special education services. There is no counterpart for nonspecial education students.

Third, special education students encompass a population of children with very diverse educational needs. With 13 federal disability categories, students in special education have a range of disabilities that include cognitive, physical, or behavioral impairments. By far the largest category, specific learning disability (SLD) accounted for over 36 percent of all special education students in the 2012–13 school year.\(^5\) As will be shown later, the average cost of educating students in special education varies considerably by category of disability and varies to an even greater extent within each category.

At the same time, the basic educational needs of most special education students are quite comparable to the needs of all students. The vast majority of special education students are held to the same academic standards,\(^5\) and consistent with LRE requirements, over one-half of the special education population spent 80 percent or more of the school day in a general education class in 2012–13.\(^5\)

### Special Education Spending

While some states have data on annual spending on special education services, the most current information across a broad array of states comes from a survey conducted by the Center of Special Education Finance (CSEF). However, these data are now quite old as 1998–99 was the latest school year in which this information was collected. Also, of the 30 states providing estimates of their special education spending at that time, only seven indicated a “high degree of confidence” in the data provided. Even the data from these seven states are marred by comparability problems because of the differing accounting conventions used across states to arrive at these estimates.\(^2\)

The most current and comprehensive estimates of special education spending for the nation as a whole are also dated. These come from the Special Education Expenditure Project (SEEP), conducted by the American Institutes for Research (AIR), and are based on data from the 1999–2000 school year. During this year, a total of $360.6 billion was spent on elementary and secondary education for all public school students in the United States. Also in 1999–2000, SEEP estimated that total spending to educate students with disabilities was $78.3 billion. This estimate includes both special and general education services for special education students, as well as other special needs programs (e.g., Title I, English language learner services, and gifted and talented education). Of the $78.3 billion, about $50 billion was spent on special education services alone; $27.3 billion on general education services; and $1 billion on other special needs programs. Put another way, special education students, who represented about 13.2 percent of all public elementary and secondary students, accounted for about 21.4 percent of the total spending on elementary and secondary education in 1999–2000. Spending on just the special education services these students received ($50 billion) accounted for 13.9 percent of spending.

The SEEP study is the last of four national special education expenditure studies that have been conducted (Rossmiller, Hale, & Frohreich, 1970; Kakalik, Furry, Thomas, & Carney, 1981; Moore, Strang, Schwartz, & Braddock, 1988; Chambers, Parrish, & Harr, 2002). Findings from all four of the national studies are summarized in Figure 32.1 in constant 1999–2000 dollars. The first of each of the four pairs of bars in this figure shows the average total expenditure (general and special education) to educate a special education student.
With the exception of the Rossmiller et al. (1970) study, it is possible to disaggregate these total estimates of spending per special education student between the general education (GE) and special education (SE) services received.

The second in each set of bars shows the average amount spent on a general education student receiving no supplemental services of any kind. For example, for 1999–2000, the first bar shows an average total expenditure per student in special education of $12,474, consisting of $4,394 for general education services and $8,080 for special education services. The second bar shows that the average expenditure on a general education student receiving no special services was $6,556.

In comparing these results, one must recognize the different methods and data sources employed by the various studies. However, based on these data, estimated real spending per special education student rose substantially from the time of the first study (1968–69) to the second (1977–78). While the 1985–86 study shows only a slight increase from the prior estimate, spending per student was considerably below the special education spending estimate from SEEP for 1999–2000.

Also of fiscal policy interest, the ratio of average spending on a special education student to spending on a general education student with no special needs increased from 1.92 in 1968–69, to 2.15 in 1977–78, to a high of 2.28 in 1985–86, and then shows a decline to 1.90 in 1999–2000. In other words, in 1999–2000, total spending on an average special education student was an estimated 90 percent more than on the average general education student receiving no supplemental educational services.

The recent decline in the spending ratio breaks the rising trend from the prior three studies. It also seems somewhat counterintuitive given that total spending on students with disabilities increased from about 16.6 percent of total education spending in 1977–78 to 21.4 percent in 1999–2000.

![Figure 32.1](https://example.com/figure321.png)

**Figure 32.1** Total Spending per Special Education Student Over Time, in Comparison to the Average Student Receiving No Supplemental Services (in constant 1999–2000 dollars).

*Source:* Adapted from Chambers, Pérez, Harr, and Shkolnik (2005).
A primary reason for the decline in the ratio between the last two studies is that increases in spending per general education student outpaced spending per special education student. Average spending per general education student increased by more than 50 percent ($4,326 to $6,556), in comparison to a 27 percent increase ($9,858 to $12,474) for special education students.

In addition, over this same period, the percentage of students aged three to 21 who were receiving special education services increased from about 11 percent to more than 13 percent of total public enrollment. This rise in the percentage of students in special education increased aggregate spending on special education. At the same time, however, it may have diminished growth in the average spending per special education student. While some of the growth in the special education population was in severe disability categories such as autism, the majority of new students were in milder and less costly categories, such as specific learning disability, speech or language impairment, and other health impairment. As shown later in this chapter, these three disability categories have the lowest spending estimates per student.

**Variations in Spending Across States**

While the national SEEP study does not have samples of sufficient size to be representative of specific states, 11 states opted for individual SEEP studies with samples large enough to generate state-specific results. Among the seven SEEP states whose results are public, total spending on the average special education student ranged from $10,141 to $15,081 (in 1999–2000 dollars), and the spending ratio ranged from a low of 1.57 in Alabama to a high of 2.55 in Maryland. This variation in the spending ratio across states reflects a combination of the overall willingness of a state to spend money on educational services, the relative priority of general versus special education, and the composition of students with respect to needs and expenditures.

While the publicly available SEEP results are limited to these seven states, variations across all states are observed in key special education personnel resources as reported to the federal government. By applying standardized salaries to the total number of special education teachers, related service providers, and aides reported by each state, one can approximate special education personnel spending per student. These estimates suggest special education personnel resources per special education student are over three times greater in the highest versus the lowest allocating states.

**Variations in Spending by Category of Disability**

As the population of special education students is quite heterogeneous—in the types of disabilities and degree of need—average expenditures mask considerable variation. The highest average total expenditures in 1999–2000 were for students with multiple disabilities ($20,095) as shown in Figure 32.2. In contrast, the two most common disabilities,
specific learning disability and speech or language impairment, which made up more than 60 percent of the special education population, show the lowest per-pupil expenditures, at $10,558 and $10,958, respectively. Thus, the expenditure for a student with a disability ranges from 1.6 (specific learning disability) to 3.1 (multiple disabilities) times the average expenditure of $6,556 for a general education student.

Disability labels are intended to reflect, to a degree, a particular student’s needs, and it is generally accepted that some disability categories are more expensive to serve than others. However, students with the same disability label may have quite different levels of severity, which translate to variations in expenditures within the same disability category. Figure 32.2 also shows the upper and lower bound of the 95 percent confidence interval around each estimate. The estimate for expenditures on students classified as specific learning disability (SLD) has a relatively narrow confidence interval ($9,807 to $11,309), whereas the confidence interval for expenditures on students with visual impairments/blindness (VI/B) is relatively wide ($15,514 to $22,108). In fact, estimates for expenditures on students with visual impairments/blindness are not statistically significantly different from expenditures on students with hearing impairments/deafness (HI/D), traumatic brain injury (TBI), autism (AUT), or multiple disabilities (MD).

The presence of large confidence intervals around some of these expenditure estimates suggests that there is a wide range of needs represented within some of these disability categories. Two students with the same disability may have very different expenditures because they have different needs and therefore receive different types and levels of services. Analysis by Chambers, Pérez, Socias, Shkolnik, Esra, and Brown (2004) shows that student disability explains only about 10 percent of the variance in per-pupil spending on individual
special education students.  

**Funding Special Education**

Special education services are primarily funded by state and local revenues. According to a state survey conducted by the Center for Special Education Finance (CSEF), federal funds in 1998–99 covered an average of 10 percent of spending on special education services, with state and local funds accounting for an average of 50 percent and 40 percent, respectively. Across the states, however, there are notable differences. The reported local share ranged from zero to 80 percent, and the state share varied from 3 to 90 percent. Given the substantial increases in federal funds in occurring in the early years of the 21st century, it is likely that this distribution has changed somewhat since 1998–99.

The formulas used by the federal government and the states respectively to distribute funds for special education differ considerably in their orientation and their detailed provisions. At the most fundamental level, they can be divided across the two broad classifications: census-based and noncensus systems.

*Census-based systems* distribute special education funds based on the total enrollment (both special and general education students) in a district (state formulas), or the total school-aged population of the state (federal formula). For example, under the most straightforward implementation of a state-level census-based funding system, districts with identical enrollments would receive the same special education aid regardless of the number of students placed in special education, the disabilities of these students, where they are placed, or how they are served. According to a 2008–09 survey by Project Forum, 10 states had adopted various forms of census-based special education funding systems (Ahearn, 2010).

*Noncensus systems*, on the other hand, often provide varying funding amounts per student based on some special education attribute, such as the overall number of special education students, the category of disability, the types of services received, and/or the primary placement for these services, for example in a general education class, a special education class, or in a separate public or private school.

A frequently cited advantage of census formulas is that, because funding is not based on the types of services provided, category of disability, or type of placement, they are free of any possible fiscal incentives to identify more students for special education (Parrish & Kaleba, 1998). At the extreme end of this argument, Greene and Forster (2002) concluded that the nation could save over $1.5 billion per year in special education spending if all states were to adopt census-type funding formulas. Although Greene and Forster’s analyses and conclusions have been challenged, incentives in special education funding with the potential to affect practice are a major policy concern (Mahittivanichcha & Parrish, 2005).

A counter-argument, however, is that such systems simply replace one set of incentives with another. For example, under census-based formulas, the fiscal incentive is to identify fewer students for special education as the revenues for these services are unrelated to the number of identified students (Parrish & Verstegen, 1994).

**Federal Funding**

The federal government converted to a census approach under the 1997 IDEA amendments. Prior to the formula change, virtually every newly identified special education student generated additional federal funds, whereas under the new formula, federal funds are allocated according to factors that are not directly related to special education: overall residential population and poverty. Policy deliberations at the time of this change suggest that it was at least partially driven by concerns about the expanding special education population.

Under the new formula, the federal government distributes 85 percent of IDEA grants to the states according to the total age-relevant residential population of each state, with and without disabilities. The remaining funds (15 percent) are allocated according to the relative degree of poverty in the state. To ease the transition to this new formula, the funding levels under the old formula in 1997 were maintained as a constant base for each state, and only the revenues in excess of $4.9 billion are distributed by the new formula. Since 2000, when federal revenues first exceeded this threshold, there have been substantial increases in federal special education funding, and, as a result, by 2013, more than half of school-age federal IDEA funds were allocated using the
new formula.

As an increasing percentage of federal special education funds are allocated under this new formula, the disparity in federal funding per special education student across states will continue to grow (Harr & Parrish, 2005). Due to differing percentages of students served in special education and varying rates of poverty, federal special education funding varied across the states from approximately $1,000 to $1,500 per special education student in 2003. Given projections that this disparity will grow substantially over time, it is possible that the issue of census-based special education funding at the federal level may be revisited in the future.

A special education finance policy issue often debated at the federal level is “full funding.” When the IDEA was first passed in 1975, it authorized the federal government to appropriate funding for each special education student up to a level equal to 40 percent of the average per-pupil expenditure (APPE) in public elementary and secondary schools in the United States. To clarify, full funding is not 40 percent of the cost of educating special education students, but rather 40 percent of the APPE for all students, including special education students. Although federal funds have increased substantially in recent years, they have never reached even half of this amount. Federal support as a percentage of APPE has steadily increased from its 1996–97 level of 7.3 percent to 15.3 percent by 2009–10, but is still considerably below “full federal funding” of 40 percent APPE.

Rising special education costs have increased the pressure for the federal government to provide additional funding up to the 40 percent of APPE, which some have interpreted as a federal commitment or promise.

State Funding

State revenues remain the primary source of special education funding, and the debate on how best to allocate these funds among local districts dominates the fiscal policy discussion. A 2008–09 survey by Project Forum showed some changes in state funding mechanisms since 1999–2000, with eight states considering future changes in the way they allocate special education funds to districts (Ahearn, 2010). Seven states reported not having a separate funding stream for special education, and the vast majority of the states used a noncensus approach to special education funding. While only seven states used a census-based approach as the primary component of their special education funding system, these included populous states such as California, Massachusetts, and Pennsylvania. Given the large size of states using this approach and the federal formula, a considerable percentage of special education revenues nationally are allocated on some form of a census approach.

The most common approach for distributing state funds to districts uses pupil funding weights, with 19 states citing such weighting as the primary basis of their special education funding approach. Pupil weights provide a basis for differentiating funding by some characteristic of the student, such as category of disability, location of primary placement, or some combination of the two. Other special education funding approaches include resource-based funding (six states) percentage reimbursement (five states), and block grants, whereby prior year’s funding is adjusted for growth in enrollments, expenditures, or inflation (one state).

Pupil weights have the advantage of recognizing cost differentials among types of special education students and generating revenue amounts that can follow individual students wherever they are served. These advantages may be particularly relevant as states increasingly move to systems of greater student and family school choice. For example, if a student with a disability chose to go to a charter school, a pupil weight system could clarify the amount of funds that should follow this child to a new school assignment (Parrish & Bitter, 2003).

A common concern in considering alternative approaches to funding special education relates to fiscal incentives. Pupil-weighted formulas that distribute state funds on the basis of some special education attribute (e.g., greater funding for certain categories of disability or placements) may encourage higher identification rates of certain disabilities or more costly placements.

In fact, the 1997 amendments to the IDEA raised concerns about higher state funding for more costly placements, which are also found to be more segregated or restrictive. States such as New York, which had a special education weighted formula based on placement, were required to show that such a formula had not resulted in failure to educate students in the least restrictive environment (LRE) appropriate to their needs as required by the IDEA. At that time, New York responded to this challenge by adding a new weight to its formula that provides more funding for serving special education students in general classes, as opposed to
There is no single best approach to funding special education. Criteria for evaluating special education funding formulas were initially developed by Hartman (1992) and expanded by Parrish (1994). These criteria describe such formula attributes as being fair, adequate, cost-based, flexible, and understandable, among others. While each of these criteria likely holds value for some constituency, the relative importance placed on each may differ. No funding formula can fully accommodate all of these criteria. As a focus on one criterion often comes at the expense of one or more of the others, the "best" formula for a given state is one that maximizes criteria that are relevant to the context of that state.

In addition to the base formula, many states have established a "high cost risk pool" to protect districts against extraordinarily high special education costs that may hit some districts harder than others in a given year. Small districts are especially vulnerable to the disproportionate fiscal impact that can result from one or two unusually high-cost special education students (Chambers, Pérez, & Esra, 2005; Richmond & Fairchild, 2013). Adjustments for high-cost students are especially important in states with census-based special education funding systems, as it is not connected in any way to true variations in special education need or cost.

Irrespective of funding approach, substantial variations in average spending per special education student can occur across districts within a given state. Given that some disability categories are more costly to serve than others, and the distribution of these high-cost disabilities is likely to vary across districts, one would expect some degree of disparity in special education spending. However, in some cases, observed spending disparities do not always appear related to cost factors such as student poverty or the percentage of students with low incidence—and generally higher-cost—disabilities.

For example, Wyoming has a special education funding system whereby the state reimburses districts for 100 percent of their eligible special education expenditures. Despite this type of funding system in which disparity in spending would seem clearly linked to differences in student need, variations in special education spending per student ranged from $7,082 in the lowest quartile of districts, to $10,382 in the highest quartile, with no apparent relationship to poverty, the percentage of students in low incidence, or higher-cost categories of disability (Parrish, Harr, Pérez, Esra, Brock, & Shkolnik, 2002).

In California, which distributes funds to Special Education Local Plan Areas (SELPAs) based on a fixed amount for each student in average daily attendance (e.g., census-based), the average special education expenditure per student in SELPAs at the 80th percentile of spending was 35 percent higher than in SELPAs at the 20th percentile of spending (Parrish, 2012). Similar spending disparities were observed in Oregon, which has a single pupil weight system (e.g., a fixed amount per special education student). Average spending per student in the quartile of the highest spending districts was over twice that of the lowest quartile (Parrish & Harr, 2007). Again, differences in poverty levels or the percentage of students identified with low incidence disabilities did not seem to drive these differentials in either state.

Other Special Education Finance Policy Issues

Other important issues are the encroachment of special education on general education revenues, the overall litigious climate surrounding special education, and the interaction between the increased focus on education accountability and how it pertains to special education.

Encroachment

The concept of special education encroachment is a growing policy concern at state and local levels of special education funding (Parrish, 2012; Asimov, 2006; Little Hoover Commission, 1997; Beales, 1993). Encroachment is generally defined as offsetting special education costs with funds that are not specifically designated for this purpose. A recent estimate of special education encroachment in California is that it exceeded $4.4 billion in the 2010–11 school year (Parrish, 2012).

With limited federal and state financial support designated for special education, and with rising special
education costs, school districts report that they are increasingly tapping into funding sources that have traditionally supported general education activities to cover the legally obligated special education services. When growth in required special education spending exceeds the state and federal revenues allocated for this purpose, the additional funds in support of special education services most generally come from the local general fund. As a result, all educational services other than special education may be curbed to accommodate required spending on special education. These program reductions may exacerbate tensions between the special education and all other public school programs and services. It also draws increasing attention to the disparity in treatment between special education and all other students, where the former has a legal entitlement to services that must be provided and the latter does not.

The issue of encroachment is complex and contentious. Synonymous with violation or intrusion, the very term “encroachment” suggests some form of hostile takeover rather than collaborative support. For example, it seems to imply that there is no local responsibility to support special education beyond the funds the federal government and the states provide specifically for special education services. In fact, what was most commonly reported by the states in the 1999–2000 CSEF survey on special education revenues and spending was a three-way support system based on some combination of federal, state, and local funds (Parrish et al., 2004). What one state calls encroachment, another may simply refer to as the expected local share. As an example, the $4.4 billion encroachment reported by California represented about 40 percent of total special education spending for 2010–11, which is similar to the 36 to 40 percent local share reported, on average, across all states in prior CSEF and federal data collections (Parrish, 2012).

Figure 32.1, shown earlier in this chapter, supports one counter-argument to the notion that special education is increasingly taking resources away from nonspecial education programs. From the time of the first expenditure study shown in this figure, which was conducted in 1968–69 prior to the implementation of federal special education law, to the most recent study using 1999–2000 data, spending per general education student increased at a greater rate than total spending per special education student. Thus, general education has realized a dramatic increase in support over a period of time during which considerable supplemental funds were also allocated to serve students in need of special education services. At the same time, rising special education costs can be seen as coming largely from the growing percentages of students being served in special education. Although the most effective way to stem the tide of rising special education spending may be to bolster general education, special education is increasingly seen as a drain against those resources. This reality undoubtedly fuels some of the frustration in regard to what is seen as special education encroachment.

One could argue that special education may be among the least efficient ways to address learning problems that could be addressed through improved general education services. This may be one reason why the reauthorization of the IDEA in 2004 included changes in how to identify children with specific learning disabilities (SLD), by far the largest category of children in special education. Representing a substantial change in special education law, the 2004 IDEA regulations moved away from sole reliance on a discrepancy model for identifying students with specific learning disabilities. This model identifies students in need of special education based on the difference between the student’s intelligence and his or her actual achievement. Often described as a “wait and see” approach, in which students can be in third or fourth grade before having a large enough discrepancy, critics of this model have argued that it can be harmful to students by potentially delaying intervention. Furthermore, many students with SLD show minimal gains in achievement after being placed in special education, and few actually exit special education (Donovan & Cross, 2002).

As an alternative to the discrepancy model, the “response to intervention” approach uses research-based educational interventions to target at-risk students within general education early, which may preclude subsequent identification for special education. Students who do not respond to increasingly intensive interventions would be subsequently identified for special education. This increased focus on remediation within general education may be one reason for the somewhat declining percentage of students identified for special education over the past several years and may also provide some relief in regard to the growing share of public education spending going to special education. Another change may lead to targeting students early and more effectively. A new provision in the 2004 IDEA allows up to 15 percent of the federal funds to be used for early intervention services for students not
identified for special education, but who need additional support to succeed in the general education environment. Flexibility of this type has the potential to benefit students who might otherwise be in special education and to bolster general education services.

**Litigation**

Concerns are frequently raised about the burgeoning costs of special education litigation arising from the legal entitlements associated with special education. The IDEA established procedural safeguards—which include due process hearings, complaint resolution, mediation, and the right to a civil trial—to resolve disputes between parents and school districts regarding the education of students with disabilities.

SEEP estimates show that districts spent $146.5 million on these due process activities (including litigation) in 1999–2000, accounting for only 0.3 percent of total special education spending in that year (Chambers, Harr, & Dhanani, 2003). If we apply this figure to all special education students, the average expenditure per special education student is $24. The rates at which these procedural activities occur are also relatively low, with a reported five due process hearings, seven mediation cases, and less than one litigation case per 10,000 students with disabilities (U.S. General Accounting Office, 2003; Chambers et al., 2003).

While the data seem to suggest that concerns about the cost of special education due process guarantees may be overstated, individual cases can sometimes be quite costly for districts. Furthermore, the SEEP cost estimates do not include the degree to which the threat of litigation may be causing districts to provide more special education services than they believe are required, thus indirectly driving up special education costs (Asimov, 2006; Nussbaum, 2003).

**Accountability, Adequacy, and Efficiency**

A final set of issues pertains to changes in the focus of accountability for special education programs and a new emphasis on Efficiency. Historically, most emphasis has been placed on fiscal and procedural accountability. More recently, however, the focus of special education accountability has shifted to educational outcomes. The federal Elementary and Secondary Education Act (ESEA), as reauthorized by the 2001 No Child Left Behind (NCLB) Act, and state-specific accountability systems clearly reinforce this conceptual shift. The ESEA specifies participation requirements for the testing of specific subgroups of the student population, including those in special education. It also requires that test scores be reported separately for each racial, economic, and programmatic subgroup within each school, and specifies clear school- and district-level sanctions when these participation or performance goals are not met.

The increased emphasis on clearly specified educational outcomes for all students has led to a number of court cases related to the adequacy of education funding (Koski & Hahnel, this volume). The resulting judicial pressure has forced many state policymakers to try to determine the cost of providing an adequate education, given the state’s implicit (or sometimes explicit) education outcome standard (Downes & Stiefel, this volume). Virtually all of the adequacy studies that have been conducted in various states incorporate special education to some degree. The attention given to special education adequacy, however, has been mixed and has been far less thorough than that given to general education (Harr, Parrish, Chambers, Levin, & Segarra, 2006).

The term “cost” in such adequacy studies is defined as the minimum expenditure required to achieve a specified standard of educational outcomes. Adequacy debates and litigation proceedings across the nation are complicated by the nature of special education law and the students it is designed to serve. For example, how do outcome standards apply to students with severe disabilities? How can the cost of adequacy be determined across the broad array and mix of services appropriate for individual special education students? In addition, the very nature of the federal law, in which every eligible student must be served in accordance with an IEP, seems counter to the notion of a standardized set of state adequacy resource standards.

The concept of the IEP arguably reflects one basis for considering adequate education for special education students. The IEP is a legal contract in which the unique needs of the student must be fully examined by a multidisciplinary team of appropriate professionals to establish educational goals and to specify the location, frequency, and duration of services necessary to make progress toward those goals. According to the federal
special education law, the cost of the services alone cannot be used as a reason to preclude services considered necessary to provide a student with disabilities with a free and appropriate education. Supplementing these IDEA provisions, the NCLB contains further requirements to ensure that professionals delivering various educational services have appropriate qualifications.

One concern is that these provisions may lead districts to provide services beyond what is needed to reach a specified education standard, and hence lead to excessive spending. While these concerns are valid, the considerable pressure regarding rising special education spending may serve as a counterbalance. In addition, special education law and regulatory provisions require audits and other checks designed to align IEPs with specified education goals and to not authorize services beyond what each child’s educational objectives require.

For these reasons, it might be argued that current practice reflects a better and more systematic assessment of what is needed to achieve specified and appropriate educational objectives for special education students than is true for other students. In this sense, studies measuring expenditures (e.g., Chambers et al., 2002, as described above) and quantities of special education staff, as reported annually by the states to the U.S. Department of Education, provide rational bases for considering special education adequacy given current levels of general education. (In addition, see the chapter by Downes and Stiefel, this volume, on alternative approaches to estimating adequacy, as well as Harr et al., 2006, which contains further discussion of special education adequacy).

Special education accountability and adequacy may also advance the concept of special education service standards. For example, what percentage of special education students might be expected to receive speech therapy in a medium-to-large school district? What is a reasonable caseload for a speech therapist? While these measures will, and should, vary to reflect differing local conditions, a greater understanding of what might be reasonably expected in regard to special education resource allocation may assist districts and states in self-evaluating their service practices and could be an integral piece of state and federal monitoring (Parrish et al., 2002; Parrish & Harr, 2005; Benzel, 2005; Levenson, 2012).

Accountability and efficiency seem especially fruitful areas for future consideration in the area of special education finance. As shown above, much of the prior research in this area has focused on inputs, i.e., the amount spent on these programs and their true costs in terms of adequate provision. Perhaps more important is the question of what is being produced as a result of these investments and how each dollar available might be used efficiently to more fully benefit the students receiving special education services (Levenson, 2012). A recent critique arguing for the need to refocus special education policy comes from Chester Finn, who wrote, “America’s approach to the education of children with disabilities is antiquated, costly, and ineffective…. The federal program is input-driven, rule-bound, compliance-obsessed, and inattentive to learning outcomes” (Richmond & Fairchild, 2013, p. 1).

What drives variations in educational, social, and emotional outcomes for students in special education and what can be learned about the factors that positively affect these outcomes? Mandated by the NCLB, comparable special education outcome information has been available across all school districts in a given state over the past decade, allowing for preliminary analyses of this nature. Studies from three states (California, Illinois, and Massachusetts) examining these states’ district-level data show dramatic variations in the percentage of students with disabilities scoring proficient or above in reading and math, even when such factors as the percentage of students in poverty, identified for special education, and breakouts by categories of disability are statistically controlled (Parrish, 2012; Parrish, 2010; Hehir, Grindal, & Eidelman, 2012). All three of these studies found a statistically positive relationship between educational outcomes and the degree to which these students were receiving educational services in general education classrooms. In addition, in the two studies where average per-pupil special education spending was also analyzed (California and Illinois), no relationship was found between this measure and educational results. These data suggest that spending more, without corresponding changes in policy, may have limited to no effect on improved outcomes.

**Conclusion**

To a large degree, students in special education are like all other students. At the same time, the degree of variation in the characteristics and educational needs within this category of students is immense. Regardless
of where they fit on the broad spectrum of disability needs, students in special education are distinct from all other students by virtue of their educational entitlement. School districts are contractually bound to provide students in special education with a full set of supplemental services appropriate to their individual needs that, according to the judgment of a professional team, will allow the students to meet specified educational outcomes.

This legal entitlement to special education services has sometimes produced tensions. Given escalating special education spending, increased outcome-based accountability for all students, and adequacy-based school finance litigation across the states, we expect increased discussion of some form of special education service standards to emerge as well as an increased focus on educational outcomes. Such standards could further the consideration of what types and quantities of special education services are adequate for students in special education, how this relates to adequacy standards for all students, and what mix of resources, policies, and procedures appear most promising in bolstering education results. Better understanding the degree to which and in what ways investments in special education actually lead to improved academic, social, and emotional results, and therefore improved life outcomes, seems an especially compelling topic for special education finance research.

Notes

3. Although over a decade old, this was the last national study conducted on special education spending.
5. In 2004–05, 98 percent of students with IEPs took statewide reading assessments, with about 12 percent being provided alternative assessments (O’Reilly, Fafard, Wagner, & Brown, 2006). This suggests that the vast majority of students with disabilities are held to the same standards as their nondisabled peers.
8. These figures include preschool special education students.
9. Estimates of per-pupil expenditure for a general education student are based on a combination of data from the SEEP school surveys and SEEP surveys for those special education students who spend the vast majority of their time in the general education classroom. Expenditures for these students include both direct instruction as well as administration and support services provided to the typical general education student.
11. The 1977–78 school year was two years after passage of the Education for All Handicapped Children Act, PL 94–142, the predecessor to the IDEA.
12. In addition to the national SEEP, all 50 states were invited to extend their participation in this study to obtain state-representative samples that could be used to address state-level policy concerns related to special education finance. Eleven states opted for a state-specific SEEP study, and 7 of the 11 states provided permission to release the results.
13. For comparability across the states, the figures presented are for school-age students only (excluding preschool).
14. Data used for this analysis included the 2004–05 counts of special education personnel and special education students as reported by the states to the federal government (http://www.ideadata.org), and national teacher salary for 2004–05 from the National Center for Education Statistics (http://nces.ed.gov/programs/digest/d05_tables/d05_077.asp). For the aide salary, we applied the ratio of special education teacher’s salary to an aide salary to the NCES teacher salary.
15. These expenditures on special education students include personnel and nonpersonnel expenditures on general education instruction, special
education instruction and related services, other special need programs (e.g., Title I, ELL), general school and district administration and support, special education program administration and support, general and special transportation services, and school facilities. Average expenditure estimates for specialized equipment are not unique to the student level and therefore may not reflect the actual expenditures for each disability type. It is expected that the estimates for disability categories with high special equipment needs are understated, and estimates for disability categories that have fewer such needs are overstated. Preschool students are not included in this data. Because of the way in which SEEP staff sampled students served in external placements, expenditures for this subset of students are not included in the expenditures by category of disability and are reported separately as a group. These students in external placements are generally served in nonpublic schools or schools operated by other public agencies. This group of students exhibited the highest average total per-pupil expenditure ($25,580).

16. Chambers et al. (2004) used a multivariate analysis that controlled for other student characteristics such as age and ethnicity and a series of community and state characteristics including district size, a geographic cost of education index, and dichotomous variables controlling for the state in which services were provided (see Exhibit 2 in Chambers et al., 2004).

17. Averages are based on 37 states that reported this information.

18. As described below, the IDEA Part B formula which allocates federal revenues to the states in support of school-age special education programs is census-based with an additional adjustment for poverty. State census-based formulas may also have adjustments to the basic concept of special education funding based on total enrollment.

19. These states include Alabama, Alaska, California, Connecticut, Idaho, Massachusetts, Missouri, Montana, North Dakota, Pennsylvania, South Dakota, and Vermont.

20. For a recent perspective on alternative approaches to funding special education, see Parrish (2014).

21. The 2004 regulations of the IDEA defines APPE as the aggregate current expenditures of all local education authorities in the 50 states and the District of Columbia, including any direct expenditures by the state for the operation of those agencies, divided by the aggregate number of children in average daily attendance. Please see §300.717 in the Federal Register (June 21, 2005) for the official regulations: http://a257.g.akamaitech.net/7/257/2422/01jan20051800/edocket.access.gpo.gov/2005/pdf/05-11804.pdf.

22. The census formula was the sole mechanism in only seven of the 10 states with some form of census-based system. The other three states combined the census system with other distribution methods, such as pupil weights.

23. The lowest-and highest-spending quartiles of districts in this study had 29 percent and 30 percent poverty, respectively. The lowest-and highest-spending quartiles were also similar in the percent of the special education population with disabilities other than specific learning disability, speech, and language impairment, and emotional disturbance, with 19 percent and 16 percent respectively.

24. An Internet search of special education encroachment produced several records posted by school districts in California, documenting the amounts of general funds going toward special education expenditures.

25. Also see Greene and Winters (2007), who discredit arguments that private placements for special education students have resulted in general education encroachment.

26. Early work on student study teams, an approach similar to the “response to intervention” model, showed reductions in referrals to special education, including reductions in inappropriate referrals to special education (Shields, Jay, Parrish, & Padilla, 1989).

27. An example of this type of work can be found in Huberman, Navo, and Parrish (2012).

References


Parrish, T., & Bitter, C. S. (2003). Special education in the City: How has the money been spent and what do we have to show for it? *Journal of Special Education Leadership, 16*(1), 34–40.


33

Resource Needs for Educating Linguistic Minority Students

RUSSELL W. RUMBERGER AND PATRICIA GÁNDARA

Introduction

Although all students are expected to meet the same challenging standards, students differ greatly in their initial social and cognitive skills when they first enter school, and in the family and community resources available to support their learning during school. As a result, some students need more support and resources in school than other students. What amount and types of resources do such students need? This chapter addresses this question regarding one of the fastest growing segments of the student population—students who come from households where a language other than English is spoken—so-called linguistic minority (LM) students.

Most linguistic minority students are not yet proficient in English when they start school. These students, referred to as English learners (ELs), require additional resources and support not only to become proficient in English, but also to learn the same academic content as English-only students. Yet while the literature on school finance has recognized the resource needs of English learners—as we illustrate below—the resource needs for all children from non-English homes have been largely overlooked in this literature. For the most part, when students are classified as fluent in English—whether this occurs when they first enter school or sometime later—their needs for specialized educational support are largely ignored, so the task of determining what their resource needs are is more complex. Throughout the chapter we attempt to clarify the distinction between the larger population of linguistic minority students and the subgroup of those students who are not proficient in English.

The Linguistic Minority Population

Linguistic minorities represent one of the largest and fastest growing segments of the school-age population in the United States. According to data from the U.S. Census, there were 12 million children, ages five to 17, who spoke a language other than English at home in 2012, representing 22 percent of the school-age population (Table 33.1). The vast majority, 72 percent, spoke Spanish. In some states the proportion of linguistic minorities was much higher; in California, for example, 46 percent of the school-age population in 2012 were linguistic minorities (U.S. Census Bureau, 2014).

Table 33.1 Linguistic Minority Population Five to 17 Years of Age in the United States, 1980–2012
Over the last 32 years, the linguistic minority population has exploded relative to the English-only population. Between 1980 and 2012, the linguistic minority population in the United States more than doubled, while the English-only population actually declined. This means that virtually all of the more than six million additional school-age children in the United States over the last 32 years were linguistic minorities.

The LM population consists of two subgroups based on their level of English proficiency: students who are not yet proficient in English are classified as English learners (ELs), while students who are proficient in English are classified as Fluent English Proficient (FEP). It is important to note, however, that there is great variability in the definition of "proficiency" and the measures and criteria used to assess it across states and even from local district to local district (Ragan & Lesaux, 2006). FEP students can be further classified based on whether they were proficient when they initially entered school (I-FEP) or whether they became proficient during school, which means they were reclassified from EL to FEP (R-FEP). In general, the proportion of LM students classified as FEP increases as students progress in school. In California, for example, only 7 percent of linguistic minorities were classified as FEP in kindergarten in 2013, whereas 70 percent were classified as FEP in eighth grade.

The process for reclassifying EL students as proficient is an important policy area, and one that has resource implications as well. A number of criteria—including English proficiency and academic achievement—are used to determine whether linguistic minorities are ELs or FEP, and these criteria differ for students’ initial classification when they first enter school (I-FEP) and subsequent reclassification (R-FEP) once they progress in school (Abedi, 2008; Gándara & Merino, 1993; Linquanti, 2001; Zehler et al., 2003). Because of the differences in the ways that students are labeled as EL and reclassified as FEP, the relative size and characteristics of both the EL and FEP groups can vary greatly from one location to another and over time. This makes identifying educational needs for both groups of students difficult. Nonetheless, as we show below, students who are reclassified as FEP continue to have needs, even if they initially meet what appear to be high standards when they are first reclassified.

### School Performance

The academic achievement of linguistic minority children, particularly those who are not yet proficient in English, lags far behind children from English-only backgrounds. For example, data from the National Assessment of Educational Progress (NAEP) reveal that test scores of EL students range between 0.83 and 1.32 standard deviations below the scores of English-only students (Table 33.2). These are considered large achievement gaps. Of course as English learners become proficient in English, they are reclassified as FEP, so at the higher grades, the proportion of ELs decreases, while the achievement gap for those remaining EL students increases.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Percent of total</td>
<td>Number</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>All speakers, age 5-17</td>
<td>47,493,975</td>
<td>55,800,084</td>
</tr>
<tr>
<td>English only</td>
<td>42,925,646</td>
<td>41,806,339</td>
</tr>
<tr>
<td>Language other than English</td>
<td>4,565,329</td>
<td>11,993,745</td>
</tr>
<tr>
<td>Speak Spanish</td>
<td>2,952,462</td>
<td>8,587,188</td>
</tr>
<tr>
<td>Speak non-Spanish language</td>
<td>1,615,867</td>
<td>3,406,557</td>
</tr>
</tbody>
</table>


Table 33.2 2013 NAEP Reading and Math Scale Scores by English Learner Status
Data also reveal that even English learners who are reclassified as FEP and who initially perform higher than English-only (EO) students in the lower grades, begin to fall behind in the secondary grades (Gándara & Rumberger, 2006, Figure 33.2). Because the population of English learners declines as more and more students are reclassified as FEP, it is important to gauge the combined performance of current ELs and former ELs (R-FEP). Such an analysis conducted with California data reveals that between grades two and 11—the period when students are tested—the achievement gap between English-only students and current/former EL students remains essentially unchanged. Of course, it is important to note that any assessment of English learners with tests administered in English only, and that were not designed for or normed on EL students, will yield less-than-valid results. Thus, test data for EL students should be interpreted with great caution. Nonetheless, important decisions are made about students and schools based on these scores.

**School Readiness**

One reason for the underachievement of linguistic minority students is that they come from more disadvantaged backgrounds than English-only students and, hence, begin school at a significant disadvantage compared to their English-speaking peers. Some indicators of this disadvantage for kindergarten students in 1998–99, based on data from the Early Childhood Longitudinal Study (ECLS), are shown in Table 33.3.

<table>
<thead>
<tr>
<th>Table 33.3 Selected Indicators of Family Background and School Readiness by Language Background, 1998–99 Kindergarteners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socioeconomic status (SES)</td>
</tr>
<tr>
<td>------------------------------</td>
</tr>
<tr>
<td>Language skills</td>
</tr>
<tr>
<td>Math achievement</td>
</tr>
</tbody>
</table>

Source: Analysis of data from the Early Children Longitudinal Study of the Kindergarten Class of 1998 (N=17,424).

Note: Indicators are standardized with national mean of zero and standard deviation of one for the entire population of kindergartners.

One indicator is socioeconomic status (SES)—a composite measure based on family income, parental education, and occupational status—that provides a useful indicator of family, human, and financial resources. The data reveal a gap of 0.38 standard deviations in SES between EO and LM students at the beginning of kindergarten. Similar gaps exist in oral and written language skills and in math achievement at the beginning of kindergarten. Linguistic minority students suffer from disparities in other areas as well. For example, linguistic minority students are less likely to participate in center-based preschool; 41 percent of linguistic minority students attended center-based preschool, compared to 59 percent of English-only students (Rumberger & Tran, 2006, Table 33.4).

**Educational Goals for Linguistic Minority Students**
Identifying the resource needs for educating linguistic minorities requires determining the educational goals for these students, and then specifying the educational programs designed to meet these goals. Ideally, these two things would be specified sequentially—that is, first the goals would be established and then the programs would be developed and implemented to meet these goals. The two activities might also be carried out by different agencies; state educational agencies, for example, might set educational goals for linguistic minority students, along with goals for other students, while local education agencies might design a variety of programs to meet these goals based on local needs and capacity.

In reality, however, the two have been inextricably linked substantively and politically for more than 30 years. A number of different government agencies have specified both goals and programs for linguistic minority students over this period. One reason is that the education of linguistic minorities, as with all students, is the province of local, state, and federal government agencies. Another reason is the education of linguistic minorities has been and continues to be a politically charged issue at the local, state, and national levels.

Over the last 40 years, federal legislation, federal courts, and state legislation have together specified five educational goals for linguistic minority students, and a number of programs to meet those goals:

1. Access to the core curriculum
2. English language proficiency
3. Native language proficiency
4. Closing the achievement gap
5. Cultural competence

Below, we provide some background on the development of these educational goals and relevant policy activities as they unfolded in three areas: (1) federal legislation, (2) federal courts, and (3) state legislation.

Federal Legislation

The education of linguistic minority students was first addressed at the federal level through provisions of Title VI of the 1964 Civil Rights Act prohibiting discrimination on the grounds of race, color, or national origin. The Equal Educational Opportunities Act of 1974 (EEOA) spelled out more explicitly that the denial of equal educational opportunity includes “the failure of an educational agency to take appropriate action to overcome language barriers that impede equal participation by its students in instructional programs” (EEOA, 1974, Sec. 1703[f], p. 136).

The U.S. Congress first passed legislation that specifically focused on the education of linguistic minority students in the Bilingual Education Act (BEA) of 1968 (Title VI of the Elementary and Secondary Education Act (ESEA) of 1965). The Act provided funds for local education agencies to develop programs to provide “meaningful and equitable access for English-language learners to the curriculum” (August & Hakuta, 1997, p. 16) without prescribing a particular program of instruction. Thus, the first legislative goal for linguistic minority students was simply access to the core curriculum with no mention of the types of instructional programs that might achieve that goal.

In the 1974 reauthorization of the BEA, however, Congress encouraged “the establishment and operation ... of education programs using bilingual education practices, techniques, and methods” (BEA, 1974, Sec. 702[a]). The legislation went on to define bilingual education as “instruction given in ... English, and, to the extent necessary to allow a child to progress effectively through the educational system, the native language” (BEA, 1974, Sec. 703[a][4][A][i]). In this legislation, the federal government established a second goal for linguistic minority students—English proficiency—and specified an instructional program to meet both this goal and the goal of access to the core curriculum—bilingual education. In this case, native language instruction was simply a temporary means for learning the core curriculum.

Subsequent reauthorizations of the BEA in 1978, 1984, and 1988 shifted the focus of instruction to achieving competence in the English language, and to the acceptance of English-only programs and not simply bilingual instruction (García & Wiese, 2002, p. 155). But in the 1984 reauthorization, the Congress also specified a third goal: “competence in English and a second language” (BEA, 1984, Sec. 703[a][5][A]), although no specific
funding was provided (García, 2005, p. 96). The 1994 reauthorization reinforced the goal of bilingual proficiency by giving "priority to applications that provided for the development of bilingual proficiency both in English and another language for all participating students" (BEA, 1994, Sec. 7116[1][1]). At the same time, the legislation reflected the influence of two other pieces of legislation—Goals 2000 and the Improving America’s Schools Act of 1994—that promoted equality and quality education for all students. For the first time, Title VI programs would help linguistic minority students "meet the same challenging State content standards and challenging State performance standards expected for all children and youth" (BEA, 1994, Sec. 7111[2][B]).

Federal policy for linguistic minority students changed once again with the reauthorization of the ESEA in 2002, the No Child Left Behind (NCLB) Act. The BEA (Title VII) was replaced with the Language Instruction for Limited English Proficient and Immigrant Students (Title III of NCLB) and the term "bilingual" was completely removed from the legislation and all federal offices and programs (Wiley & Wright, 2004, p. 155). The stated purpose of Title III was "to ensure that children who are limited English proficient, including immigrant children and youth, attain English proficiency" (NCLB, 2002, Title III, Sec. 3102), although the law does allow "instructional use of both English and a child’s native language." The legislation requires developing and attaining "annual measurable achievement objectives" (AMO’s) for English-language proficiency that some critics claim are "strict, complex, and questionable" (Wiley & Wright, 2004, p. 157). The legislation also requires schools and districts to demonstrate "adequate yearly progress" for all students, including EL students, to meet state proficiency standards, with a requirement that all students reach proficiency by the year 2014. The requirement that schools must demonstrate that EL students are making adequate progress meeting the same standards required of native-born English-speaking students, in spite of significantly greater challenges and without specifying additional resource needs to meet them, placed many schools in jeopardy of losing their accreditation, sometimes in spite of their best efforts. Some critics have also called attention to the ways in which such policies can stigmatize EL students as a source of problems for their schools (Novak & Fuller, 2003). NCLB also established a fourth goal for linguistic minority students: "... closing the achievement gap between high-and low-performing children, especially the achievement gaps between minority and nonminority students, and between disadvantaged children and their more advantaged peers" (NCLB, Title 1, Sec. 1001). Many attempts have been made over the last several years to reauthorize the ESEA and to modify the federal rules, as it became increasingly clear that the goals were unattainable with existing resources. However these efforts have not met with success and most states have sought waivers from the federal rules.

**Federal Courts**

The federal courts have a long history of involvement with the education of linguistic minority students (see García, 2005, pp. 77–85). The courts have primarily focused on establishing the rights of linguistic minority students and, in doing so, also specified educational goals for these students.

The most noteworthy court case was *Lau v. Nichols* (414 U.S. 563 (1974)), which was filed on behalf of Chinese-speaking students in San Francisco who received instruction only in English without any additional support. Rejecting a lower court ruling, the U.S. Supreme Court ruled "Basic English skills are at the very core of what these public schools teach. Imposition of a requirement that, before a child can effectively participate in the educational program, he must already have acquired those basic skills is to make a mockery of public education." Essentially, the Court established the first goal—access to the core curriculum—at the same time the U.S. Congress was specifying the same goal in the BEA of 1974. Yet, like Congress, the Court did not specify a remedy, or what kind of instructional program would provide students with access to the core curriculum.

However, in 1981 the federal court in a case entitled *Casteñeda v. Pickard* (648 F.2d 989) provided guidelines for what could constitute a legitimate program for ELs: (1) it had to be based on sound theory; (2) it had to be implemented with fidelity—including adequate resources; and (3) over time it had to demonstrate that it was effective. Subsequent court cases have tended to weaken the legal underpinnings of the *Lau* decision, and thus have afforded few legal avenues to broaden or even maintain reasonable educational goals for linguistic minority students (Moran, 2004; García, 2005). In 2009 the Supreme Court delivered its second most important decision directly affecting the education of English learners. In a case entitled *Horne v. Flores* (129 S. Ct. 2579),
the Court ruled that lack of spending should not be a criterion for determining if a program is adequate and found a required Arizona program that taught only English and denied ELs access to the regular curriculum for an indefinite period of time to be legal. Thus both the EEOA and the Lau decision were stripped of much of their force.

State Legislation

The history of state legislation regarding the education of linguistic minority students is varied, in part, because many states have had relatively few linguistic minorities until quite recently. All states are required to identify their English learners and ensure that they are provided with “appropriate” services. There is, however, great variability in the options offered, the preparation and certification of teachers for these students, whether instruction is provided in native language, and to what extent the states monitor their programs (Colorín Colorado, 2011; Zehler et al., 2003). By and large, most state legislation has focused on the second educational goal to “as effectively and efficiently as possible, ... develop in each child fluency in English” (California Education Code, 1976, Section 52161).

Educational Programs for Linguistic Minority Students

Just as there are a number of different goals for educating linguistic minority students, a number of educational programs have been developed to meet these goals. Most educational programs have been developed to serve only LM students who have been identified as ELs. Most of these programs contain a number of specific elements that depend not only on the goals or standards of the program, but also on the characteristics of the EL population at the school and on the school and district context. Salient characteristics of the population that would affect resource needs include family background (income and SES), the number of years in the United States, age and grade level of the students, prior schooling, native language proficiency, initial English language proficiency, and academic achievement. For example, poor EL students with low initial levels of English proficiency would generally need more intensive programs and resources than nonpoor EL students with high initial levels of English proficiency (see NCSL, 2005). Schools with EL students from a single language background, such as Spanish, may need fewer program elements and resources than schools with EL students from many language backgrounds (Rumberger, Gandara, & Merino, 2006). Of course, the programs and services that English learners actually receive is largely unknown. At best, states collect general program information on such features as language of instruction, and there are some infrequent national surveys. In the following sections, we provide an overview of what we do know about instructional and supplemental programs for English learners.

Instructional Programs

The key element of any educational program for English learners is an instructional program that addresses at least two goals: English language proficiency and access to the core curriculum. The defining feature of the instructional program, and the one that has been the focus of most of the policy debates over the last 40 years, is the language of instruction. Some instructional programs for EL students are conducted only in English, while others are conducted in both English and the children’s native language, which are commonly referred to as bilingual programs.

Bilingual programs can be further characterized by the extent and goals of the native language instruction (see Table 33.4); in some programs, known as early-exit or transitional bilingual programs, the extent of native language instruction is limited to a relatively short time (one to three years) and the only goals are to facilitate the learning of English and provide access to the core curriculum during that transition period (which can, and often does, imply early literacy as well). In other programs, known as late-exit or developmental programs, the goal of the program is to develop grade level literacy in the native language as well as in English, so the
duration of primary language instruction is much longer, usually until the transition to middle school. Still another type of bilingual program, known as a dual-immersion or two-way bilingual program, enrolls native English speakers along with English learners with the goal of full literacy in both languages and subject matter proficiency for both groups of students.

Table 33.4 Major Program Models for English Learners

<table>
<thead>
<tr>
<th>Program</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>English immersion (&quot;sheltered English&quot;)</td>
<td>Use of English-only strategies; English as a second language; specially designed academic instruction in English; usually lasts 1–3 years</td>
</tr>
<tr>
<td>Transitional bilingual education/early exit</td>
<td>Instruction using some primary language for transitional support; may teach early reading in primary language but normally makes transition to English-only within 2–3 years; goal is literacy in English</td>
</tr>
<tr>
<td>Developmental bilingual/late exit</td>
<td>Emphasis on developing competence in two languages; students may receive instruction in primary and second language for many years; goal is literacy in two languages</td>
</tr>
<tr>
<td>Two-way immersion (&quot;dual immersion&quot;)</td>
<td>English speakers and English learners taught together with focus on two languages; emphasis on strengthening minority language; goal is biliteracy; usually lasts 6 years or more</td>
</tr>
</tbody>
</table>

As discussed earlier, the choice of instructional programs has often been influenced and sometimes prescribed by federal and state policy. And, increasingly, those prescribed instructional programs use only English. Yet an ever-growing body of research supports the value of native language instruction. The most recent and extensive review of the research by the federal government-appointed National Literacy Panel on Language-Minority Youth and Children (August & Shanahan, 2006) concluded: “Studies that compare bilingual instruction with English-only instruction demonstrate that language-minority students instructed in their native language as well as in English perform better, on average, on measures of English reading proficiency than language-minority students instructed only in English” (p. 5).

Despite the growing research base supporting native language instruction, most linguistic minority students in the United States are instructed in English. The latest national survey of school LEP coordinators found that 60 percent of LEP students were taught entirely in English, 20 percent were taught with some (2 to 24 percent) native language instruction, and 20 percent were taught with 25 percent native language instruction (Zehler et al., 2003, p. 36).

A second key feature of the instructional program is the extent it provides instructional services that are specifically designed for English learners. The national survey also found that 12 percent of EL (LEP) students received no specific instructional services, 36 percent received supplemental instructional services (e.g., aides, resource teachers, and/or fewer than 10 hours a week of ESL instruction), and 60 percent received extensive instructional services where a significant amount of instruction was designed for LEP students (Zehler et al., 2003, p. 36).

A third key feature of the instructional program concerns instructional time. To acquire English and to master the core curriculum, English learners need additional instructional time. In 1963 psychologist John Carroll developed a model of school learning that posited learning as a function of: (1) aptitude, or the time students need to learn (which is inversely related to the quality of instruction and students’ ability to understand instruction); (2) opportunity to learn, or the time allowed for learning, and (3) perseverance, or the amount of time students are willing to spend on learning (Carroll, 1989). The model suggests that some students need more time for learning than other students and that increased time allocated to learning can increase achievement (all else being equal). A number of case studies have documented how effective schools often devote more time to student learning than typical schools (Gándara, 2000; Farbman & Kaplan, 2005) and some experimental evidence has documented that increased learning time can improve student achievement (Gándara & Fish, 1994; Patall, Cooper, & Allen, 2010). Nonetheless, resources for additional instructional time for English learners are not routinely built into school budgets, and schools that attempt to meet this need often have to seek funding from other sources (Gándara & Rumberger, 2006).
A final key feature of the instructional program concerns class size. Two recent reviews of research literature found that small classes generally improve student achievement, although the impact varies in a nonlinear fashion (Ehrenberg, Brewer, Gamoran, & Willms, 2001; Krueger, 2003). One of the largest and most widely studied experimental studies was in Tennessee, where classes were reduced from 22–27 students to 13–17 students (Mosteller, 1995; Finn & Achilles, 1999). Several studies have documented both short-term and long-term benefits of small classes in Tennessee, especially for minority students (Finn, Gerber, Achilles, & Boyd-Zaharias, 2001). Despite the commonsense notions that increased instructional time and reduced class size should also affect the learning outcomes for English learners—and the fact that effective schools for these students often include such modifications (Gándara & Rumberger, 2006)—no study has ever been conducted to demonstrate these effects.

**Supplemental Programs**

Meeting the educational goals for ELs may require more than an instructional program, even an extensive program based on the primary language. English learners, especially those who come from socioeconomically disadvantaged backgrounds, may require other programmatic services beyond those found in the regular classroom during the regular school day. Such supplemental programs may include preschool, student support (such as peer or adult tutoring), and family supports. Language minority students who become proficient in English may need continued support to match the progress of their English-background peers and even more support to reach the goal of closing the achievement gap. Of course, other disadvantaged students who come from English-speaking backgrounds may also need more instructional time, smaller classes, and supplemental programs to meet the same academic standards as more socioeconomically advantaged students.

**Resource Requirements of Educational Programs**

To provide effective educational programs for linguistic minority students requires both the right amount and the right type of resources. This section discusses three critical resources needed to educate linguistic minority students: teachers, instructional materials, and assessments. The last section of the paper examines the extent to which the type and amount of resources needed to educate linguistic minority students differs from those required to educate other students.

**Teachers**

The most important resource for educating students is teachers, both classroom teachers and support teachers. While it is critical to provide a sufficient number of teachers to meet the educational goals for linguistic minority students, it is also critical to provide teachers with the proper skills, abilities, and attitudes for dealing with this population. Unfortunately, there is little rigorous empirical evidence on what attributes of teachers are most effective in improving educational outcomes for students generally, let alone for linguistic minority students (Wayne & Youngs, 2003).

But research on effective instructional approaches to educating linguistic minority students does suggest that teachers of such students need several critical competencies. First, teachers of EL and linguistic minority students need specific pedagogical and discipline-specific knowledge. Wong-Fillmore and Snow (2000) argue that all teachers, but especially those who teach EL students, need to know a great deal about the structure of language, its development in the first and second languages, and how to support and enhance it. Understanding how to use cognates in building new vocabulary, using speech markers, and frequent checks for comprehension, are skills that teachers of EL students must have to be effective (August & Shanahan, 2006). Second, teachers of English learners also need to know how to use assessments to measure language proficiency and to monitor student progress (Genesee, Lindholm-Leary, Saunders, & Christian, 2006, pp. 136–139). Third, teachers of EL students should be bilingual. Clearly, teachers instructing in the primary language
must be bilingual. But bilingual skills are also useful in English-only classrooms to provide “support” for bilingual students in their native language by previewing and reviewing instructional activities, monitoring their understanding, and motivating students by building rapport with them. And bilingual skills are invaluable for teachers to communicate and work with the parents of linguistic minority students (Gándara & Maxwell-Jolly, 2005; Hopkins, 2013). Finally, while disciplinary knowledge and pedagogical skill are important, noncognitive skills may be the most critical characteristics of effective teachers of EL students—compassion, understanding of the challenges that students face, a strong belief in the students’ natural abilities, a deep desire to see students succeed, the ability to motivate students, and a willingness to adapt their instruction to meet the distinctive needs of EL students. A recent review of this literature (Gándara & Maxwell-Jolly, 2005) confirms this view, finding that affective characteristics and “active” teaching behaviors were cited more frequently by researchers as characteristics of effective teachers than other types of knowledge.

Existing data find that relatively few teachers of English learners possess even some of these competencies. A national survey of coordinators and teachers in programs for English learners found that only 40 percent of teachers had “significant EL (LEP) training,” which the authors defined as having a bilingual or English as a Second Language (ESL) certification, or having received 20 hours or more of in-service training related to the teaching of English learners (Zehler et al., 2003, p. 72). In fact, the study found that 40 percent of all teachers of English learners had received no in-service training in the previous five years. Once English Learners are transitioned into mainstream classes, relatively few teachers are bilingual or have specialized knowledge of how to teach them.

**Appropriate Instructional Materials**

To provide a comprehensive and appropriate instructional program requires appropriate instructional materials. In addition to the core instructional materials, linguistic minority students need both strong English Language Development materials and texts, and bridging material that allows them to access the core curriculum with their more limited vocabularies and knowledge of English language structure (these may include primary language materials, as appropriate). Beyond that, linguistic minority students would benefit from computer-based instructional materials; for example, digitalized curriculum (and the hardware to run it) allows students to devote more time to studying subjects outside regular classroom hours. This can be especially beneficial for secondary students who may have little time to catch up for graduation. Bilingual materials for secondary students who must learn English rapidly and simultaneously stay at grade level in their classes are an important resource, though rarely available. Project SOL was created to provide such materials online so that students whose English was not strong enough to understand complex texts in science and mathematics could access the material in a language they understand (see Gándara, 2013). Library books need to cover a wide range of levels and be available in the languages of the students, both to stimulate reading among those who are not yet fluent in English, and to provide the opportunity for non-English speaking parents to read with their children.

National data show that relatively few EL students receive appropriate materials. Only 57 percent of EL district coordinators in a national survey reported that EL teachers in their districts were provided curriculum materials for their EL students to help them align their instruction to state standards (Zehler et al., 2003, p. 62). A study in California found that even when specifically designated materials were developed by textbook publishers, in response to state guidelines, they were often difficult for teachers to use and therefore ineffective (Calfee, 2006).

**Valid, Comprehensive Assessments**

To provide an adequate education for linguistic minority students requires valid diagnostic, formative, and summative assessments. Diagnostic assessments are needed to evaluate the skills and abilities of linguistic minority students when they first enter school, in both their primary language and English; formative assessments are needed to provide teachers with ongoing information on the progress of linguistic minority students in both language development and subject matter competence; and summative assessments are
needed to measure the progress of linguistic minority students in reaching standards and other outcomes, including noncognitive outcomes. Summative assessments are also useful for holding school systems accountable for providing adequate educational opportunities for students. Moreover, accurate assessment of students’ skills can accelerate students’ learning, allowing teachers to build on what students already know (Darling-Hammond, Ancess, & Falk, 1995), rather than assuming that students do not know material they cannot demonstrate on a test. This can be especially critical in upper grades where students who have received a formal education in other settings may be placed in courses that are below their skill level, unnecessarily handicapping them in meeting academic requirements.

The development of appropriate assessments is a more technically complex and expensive undertaking than the development of textbooks and other curricula (Abedi, 2002). These issues have been long debated within the testing industry, but without a sufficiently large market and with fluctuating state and national testing policies, there has been little incentive to tackle them. Given the numbers of English learners and linguistic minority students, with the proper signals to the test manufacturers, there would undoubtedly be interest in engaging in this test development. For example, the enforcement of NCLB’s guidelines that states use valid and reliable tests for EL students, would provide important incentives for test makers to begin developing better assessment tools for EL students. This issue has become even more complex and difficult with the advent of the Common Core Standards. The standards, like most other educational innovations, were not developed with English learners in mind, and meager funding to adapt assessments for them only came as an afterthought. Many in the EL community worry greatly that the inability of teachers to appropriately adapt the Common Core for ELs because of lack of professional development, and the invalid testing of these students, will create even greater achievement gaps (Linquanti, 2011; Ravitch, 2014).

Review of Cost Studies

The previous discussion identified a number of resources needed to educate linguistic minority students, but what is the cost of providing those resources? One way to address this question is to examine results from some recent cost studies. A growing number of states are attempting to define an adequate education and provide the resources for schools to provide it. (See Downes and Stiefel, this volume, for a discussion of these studies.) As a result, scholars have undertaken studies to estimate the costs of providing an adequate education, including the differential costs associated with providing an adequate education to students who are disadvantaged due to poverty, language background, and disability. A recent review identified 70 empirical studies published since 1990 that examine per-pupil spending for English learners (Jimenez-Castellanos & Topper, 2012). These studies are based on four different costing out methods, each with strengths and weaknesses, with 34 percent of the studies using multiple methods. One method, cost function analysis (CFA), estimates costs from a production function that includes student performance measures, pupil characteristics (such as poverty, EL status, and disability), educational inputs (such as teacher salaries), and geographic cost differences (see Duncombe, Nguyen-Hoang, & Yinger, in this volume). Another common method, known as Professional Judgment Panels or PJPs, estimates costs based on educational programs specified by panels of “expert” educators (superintendents, principals, teachers, resource specialists). Two other methods are successful school models (SSM) and evidenced-based (EB) models. Despite the large number of studies that have examined resource needs for educating English learners, relatively few have estimated the extra costs of educating English learners and none has considered the costs of educating the larger population of linguistic minorities.

Fiscal Resources

All the methods used to conduct cost studies can be used to generate overall per-pupil cost estimates and estimates of per-pupil weights associated with poverty, EL, and special education status. Per-pupil weights generally range from 0.1, which denotes a 10 percent differential cost, to 1.0 or more, which denotes that the costs are twice as high or higher to educate disadvantaged students. One important question is how the
weights associated with poverty compare to the weights associated with EL status, independent of poverty. That is, how do the additional costs of educating poor children compare with the additional costs of educating English learners?

Estimated weights from four studies are shown in Table 33.5. The first New York study (Duncombe & Yinger, 2005), using a production function approach, estimated a range of weights based on different techniques for making comparisons among districts and using different measures of poverty. Using Census data on poverty, for example, it produced estimates of poverty weights ranging from 1.22 to 1.59 (when special education students are included). These estimates suggest that to educate poor students to the same standard as nonpoor students requires per-pupil funding levels from 122 to 159 percent higher. Educating EL students, controlling for poverty and special education, requires funding levels from 101 to 142 percent higher still. Using subsidized school lunch data (which is a broader measure that includes both poor and low-income students) produced higher poverty weights—ranging from 1.36 to 2.15—but no additional EL weights. In other words, the second set of estimates suggests there are no incremental costs of educating EL students beyond the costs of educating low-income students.

Table 33.5 Estimates of Per-Pupil Spending and Weights for Poverty and EL Status

<table>
<thead>
<tr>
<th></th>
<th>Base</th>
<th>Poverty weight</th>
<th>EL weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>New York (Duncombe &amp; Yinger, 2005)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimate using Census poverty</td>
<td>1.22–1.59</td>
<td>1.01–1.42</td>
<td></td>
</tr>
<tr>
<td>Estimate using subsidized lunch</td>
<td>1.36–2.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>New York (Chambers et al., 2005)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school</td>
<td>$10,072</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>Middle school</td>
<td>$9,899</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>$10,443</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td><strong>Arizona (NCSL, 2005)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school</td>
<td>$4,198</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>Low need ELs</td>
<td>$4,195</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>High need ELs</td>
<td>$4,060</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>Middle school</td>
<td>$4,069</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>Low need ELs</td>
<td>$4,214</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>High need ELs</td>
<td>$4,127</td>
<td>.48</td>
<td></td>
</tr>
<tr>
<td><strong>California (Chambers et al., 2006)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base model</td>
<td>$10,315</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>High poverty</td>
<td>$11,502</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>High poverty, high EL</td>
<td>$12,978</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>Panel 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base model</td>
<td>$8,920</td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>High poverty</td>
<td>$12,023</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>High poverty, high EL</td>
<td>$12,215</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Middle school</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base model</td>
<td>$8,905</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>High poverty</td>
<td>$9,793</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The second New York study (Chambers et al., 2004), based on Professional Judgment Panels (PJP), estimated per-pupil weights for EL students that ranged from 0.18 for elementary schools to 0.20 for middle schools. It should be pointed out that the base funding level in this state was much higher than the national funding level of $7,904 in 2001–02 (National Center for Education Statistics, 2000b, Table 162), which may reduce the need for extra expenditures for educating poor and EL students. Yet it leaves open the question of whether uniformly higher spending has any effect on closing achievement gaps.

The third study (National Conference of State Legislatures, 2005) used the PJP approach to estimate per-pupil weights for educating two types of EL students in Arizona: low need EL students, defined by the panel as either poor with high English proficiency or non-poor with middle and high English proficiency; and high need EL students, defined by the panel as either poor with low or medium English proficiency or nonpoor with low English proficiency. The per-pupil weights for low-need EL students range from 0.24 to 0.30, while the per-pupil weights for high-need EL students range from 0.38 to 0.57.

The last study (Chambers, Levin, & DeLancey, 2006) used two separate PJPs to generate two independent estimates of per-pupil funding levels for three school configurations in California: (1) a base model that represents the “average” school configuration for an elementary, middle, and high school in the state; (2) high-poverty schools with a higher concentration of poor (free-and reduced-lunch) students, but with the same concentration of EL and special education students; and (3) high-poverty, high EL schools with both a higher concentration of poor and EL students, but the same concentration of special education students. From the first and second estimates, it is possible to compute a per-pupil weight for poverty, and from the second and third estimates it is possible to compute a per-pupil weight for EL status.

The results show that the poverty weights are mostly—and in some cases much—higher than the EL weights. The results also show very disparate estimates from the two panels. At the elementary level, for example, the first panel estimated virtually identical weights of 0.38 for poverty and EL status, whereas the second panel estimated a per-pupil weight for poverty of 1.07 and a per-pupil weight for EL status of only 0.05. The estimated costs per student in the base models differed as well. This large discrepancy in estimated costs reflects one of the criticisms of the PJP approach—that it can generate very different estimates of the costs of providing an adequate education, probably owing to very different knowledge and experience of the panel members.

### Material Resources

One of the benefits of the PJP approach is that the panel first identifies the elements of an educational program and then determines the material resources to provide it. The panels are not required to identify specific programs, simply alternative features of a basic program in terms of resources. For example, the panel may decide that the best way to provide an adequate education to more disadvantaged students is to reduce class sizes for those students or to extend learning time by either lengthening the school day or the school year.

Differences in the resource allocation decisions are illustrated by comparing the two PJP results for the California study (Chambers, Levin, & DeLancey, 2006). The material resources allocated to the three

<table>
<thead>
<tr>
<th>High poverty, high EL</th>
<th>$10,243</th>
<th>.18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base model</td>
<td>$7,899</td>
<td></td>
</tr>
<tr>
<td>High poverty</td>
<td>$10,179</td>
<td>.85</td>
</tr>
<tr>
<td>High poverty, high EL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base model</td>
<td>$9,285</td>
<td></td>
</tr>
<tr>
<td>High poverty</td>
<td>$9,890</td>
<td>.17</td>
</tr>
<tr>
<td>High poverty, high EL</td>
<td>$10,060</td>
<td>.08</td>
</tr>
<tr>
<td>Panel 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base model</td>
<td>$7,035</td>
<td></td>
</tr>
<tr>
<td>High poverty</td>
<td>$9,352</td>
<td>.87</td>
</tr>
<tr>
<td>High poverty, high EL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
configurations of elementary schools by the two panels are shown in Table 33.6. The first panel specified a class size of 20 and hence the same number of classroom teachers for all grades in all three configurations. However, they specified a higher number of support teachers and instructional aides in the high-poverty and high-poverty/high-EL configurations. They specified a seven-hour school day for all students, extending it to eight hours for all disadvantaged (poor) students. They also specified a 190-day school year for all students and another 10 days per year for teacher planning and coordination. Finally, they specified a small number of disadvantaged students to receive preschool and early childhood programs. The per-pupil weight for ELs was 0.38 in the first panel because the panel made substantial increases in the instructional personnel beyond those specified for a high poverty school, primarily for additional support teachers and teacher aides.

The second panel allocated resources much differently than the first panel, specifying larger classes for grades four through five than K–3, and fewer instructional and noninstructional personnel. They specified a shorter school day (6.5 hours) than the first panel, but a longer school year (200 days) for most students, with a smaller number of disadvantaged students having a longer school day than was specified by the first panel. Yet they also specified that more students would participate in preschool and early childhood programs. The EL per-pupil weight was only 0.05, because the panel did not specify any additional instructional personnel beyond those specified for a high-poverty school.

### Table 33.6 Resource Allocations for Elementary Schools, California Professional Judgment Panels

<table>
<thead>
<tr>
<th></th>
<th>Panel 1</th>
<th>Panel 2</th>
<th>Panel 1</th>
<th>Panel 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base</td>
<td>High poverty</td>
<td>High poverty/ high-EL</td>
<td>Base</td>
</tr>
<tr>
<td>Enrollment</td>
<td>516</td>
<td>516</td>
<td>516</td>
<td>516</td>
</tr>
<tr>
<td>Percent year</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>Percent EL</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>Percent Spanish ELs</td>
<td>79</td>
<td>79</td>
<td>79</td>
<td>79</td>
</tr>
<tr>
<td>Percent special education</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>1. Class size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K–3</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>4–5</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>2. Personnel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom teachers</td>
<td>25.5</td>
<td>25.5</td>
<td>25.5</td>
<td>24.0</td>
</tr>
<tr>
<td>Support teachers</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
<td>8.0</td>
</tr>
<tr>
<td>Instructional aides</td>
<td>6.2</td>
<td>6.2</td>
<td>6.2</td>
<td>4.8</td>
</tr>
<tr>
<td>Substitute teachers</td>
<td>1.6</td>
<td>1.6</td>
<td>1.6</td>
<td>1.2</td>
</tr>
<tr>
<td>Total instructional personnel</td>
<td>39.3</td>
<td>39.3</td>
<td>39.3</td>
<td>36.5</td>
</tr>
<tr>
<td>Instructional and pupil support personnel</td>
<td>12.0</td>
<td>12.0</td>
<td>12.0</td>
<td>12.0</td>
</tr>
<tr>
<td>School administration</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>4.9</td>
</tr>
<tr>
<td>Maintenance and operations</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Total noninstructional personnel</td>
<td>19.0</td>
<td>19.0</td>
<td>19.0</td>
<td>19.0</td>
</tr>
<tr>
<td>Professional development (days/year)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3. School day (hours)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruction</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
<td>6.5</td>
</tr>
<tr>
<td>Disadvantaged students</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>Number of students served (95)</td>
<td>190</td>
<td>190</td>
<td>190</td>
<td>190</td>
</tr>
<tr>
<td>Teacher planning and coordination</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>4. School year (days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instruction</td>
<td>190</td>
<td>190</td>
<td>190</td>
<td>200</td>
</tr>
<tr>
<td>Disadvantaged students</td>
<td>190</td>
<td>190</td>
<td>190</td>
<td>200</td>
</tr>
<tr>
<td>Teacher planning and coordination</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>5. Early childhood program for 3-year-olds (95)</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>85</td>
</tr>
<tr>
<td>6. Early childhood program for 4-year-olds (95)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>65</td>
</tr>
</tbody>
</table>


All of these approaches specified the same level of adequacy for all students, and thus did not consider different standards for EL students as discussed above. Achieving higher standards will undoubtedly require additional resources (Gándara & Rumberger, 2008).

### Summary of Findings

The cost studies reviewed above provide little consensus on either the amount or types of additional resources...
that are needed to educate English learners above and beyond those needed for other disadvantaged students, particularly poor and low-income students. The recent review of 70 cost studies reached a similar conclusion (Jimenez-Castellanos & Topper, 2012). This appears to depend, in part, on how the population of economically disadvantaged students is measured and, consequently, the size of the population. Estimates based on a narrower definition that includes only students living in poverty (Duncombe & Yinger, 2005, Table 6) suggest that additional resources may be needed to educate English learners because many such students are not poor (see Table 33.4). However, estimates based on a broader definition that includes both poverty and low-income students (as captured by students enrolled in the federal school lunch program), suggest that no additional resources may be needed. But even when a broader and more inclusive definition of economically disadvantaged is used, a lack of consensus remains (Chambers, Levin, & DeLancey, 2006).

Additional Resource Needs for Linguistic Minority Students

A critical point of discussion in a study of resources for linguistic minority students is the degree to which their needs differ from, or are in addition to, the needs of both all other students and other socioeconomically disadvantaged students. A variety of evidence suggests that all students need qualified teachers, additional support personnel, appropriate instructional materials, and sufficient instructional time to meet high performance standards; all teachers should possess an array of human resources to make them effective, including subject matter knowledge, pedagogical skills, multicultural skills, empathy, efficacy, and the willingness to learn and work collaboratively (Banks, 1988; Cohen, Raudenbush, & Ball, 2003; Sleeter, 2001); and all schools should possess social resources that foster close and caring relationships among teachers, students, parents, and administrators (Ancess, 2003; Barnes, 2002; Bryk & Schneider, 2002; Goodard, Hoy, & Hoy, 2000; Spillane, 2004). Virtually all cost studies and studies of effective schools also acknowledge that there are additional resources required to educate all disadvantaged students who begin school with lower academic skills relative to their more advantaged peers, and thus have more to learn to reach the same achievement standards. Perhaps the most critical resource for schools educating disadvantaged students, above and beyond qualified teachers, is more instructional time, because those students have more to learn to reach the same standards as students who arrive at school with expected initial achievement levels.

What is more difficult is to determine whether the resource needs to educate linguistic minority students—both those who are not proficient in English (ELs) and those who are (FEP)—are similar to those required for poor and low-income students. We believe the evidence suggests that some needs of linguistic minorities are indeed different from other students with similar socioeconomic backgrounds and their needs cannot all be met with the same set of resources; however, it is not clear to what extent—if at all—they require more resources than those of poor and low-income children. Data from a cohort of children who entered kindergarten in the fall of 1998 indicate that the initial achievement gaps are larger between poor and non-poor students than they are between linguistic minority and English-only students, which suggests that more resources are needed for poor than linguistic minority children (Table 33.7). The data also show that students who are both poor and linguistic minority have the largest gaps in initial achievement, which suggests they may need more resources than students who are either poor or linguistic minority. Yet by fifth grade, achievement levels for linguistic minority students show greater improvement than achievement levels for poor students. Moreover, achievement levels for poor, linguistic minority students were either similar (in the case of reading) or higher (in the cases of language and math) than those of poor students from English-only backgrounds. These data suggest that achievement gaps due to language are more amenable to intervention and therefore may require fewer additional resources than the gaps created by poverty.12

Table 33.7 Selected Background and Achievement Measures by Poverty and Linguistic Minority Status, Fall 1998 Kindergarteners

515
Many resource needs appear to be similar for both poor and linguistic minority students, such as access to teachers who are skilled in strategies for developing their linguistic and academic skills, and to counselors and other ancillary personnel who can effectively communicate with their parents. But some resource needs are unique to linguistic minority students. In general, these are resources that use the students’ primary language and are created with language difference in mind. Students in primary language programs require instructional materials in the students’ native language, and all teachers of linguistic minority students, even those in English-only programs, should have access to materials in the students’ home languages, to the extent practicable, so that parents who do not speak English can be enlisted in supporting their children’s learning. Other material resources include assessments in the students’ primary language, or other assessments that can more accurately test their skills, and English Language Development materials that are designed for non-English speakers. The other unique resource consists of teachers and staff who speak the languages of the students. To recruit and attract bilingual teachers may require financial incentives, especially in communities experiencing rapid increases in linguistic minority populations. Some communities have been able to recruit and train teachers from the ranks of bilingual paraprofessionals (Flores, Keehn, Perez, 2002). Any additional costs for bilingual teachers may be mitigated by decreased use of bilingual classroom aides and other pull-out programs. For example, research has shown that providing EL students with a bilingual teacher, as opposed to relying on additional support personnel to augment the instruction of monolingual English-speaking teachers, is the less costly alternative and may result in no additional expenditures (Carpenter-Huffman & Samulon, 1981; Parrish, 1994).

In sum, English learners and other linguistic minority students, do require additional resources, above and beyond those of all other students, but their needs appear to differ more in kind than in quantity from those of poor and low-income students who are also struggling with developing broader vocabularies, a command of academic English, and familiarity with the cultural capital that are such important academic assets for the middle class.

Conclusions

The growing population of linguistic minority and English learner students in the nation’s schools has significant implications for the funding and designing of effective educational programs to educate these students. It also has implications for how all students are educated in contexts in which linguistic minority students sit alongside English-speaking peers in the classroom. Both the educational goals for LM students, and whether the programs to educate them make use of primary language instruction, have been subject to considerable policy debate over the last 30 years, even while a growing body of research evidence documents the benefits of primary language instruction for both English learners and English speakers, when it provides the opportunity for the latter to acquire a second language. Less clear from the research literature are the types and amount of resources needed to educate linguistic minority and English learner students, and the extent to which these resources differ from those needed to educate all other students and other disadvantaged students, particularly poor and low-income students. The research does suggest that additional resources are needed to
educate all students to higher standards, but to bring linguistic minority students—and others who enter school substantially behind their English-speaking classmates—to the level of their more advantaged peers requires still more educational resources, including: suitably qualified teachers; appropriate instructional materials; valid, comprehensive assessments; and probably additional instructional time. As the states continue to struggle with adequate educational funding to help students meet high standards, it is critical that they take into account the specific needs of English learners and other linguistic minorities. While more research will be needed to determine the exact additional costs for these students, if any, it is clear that their needs are, at least in part, different from their native-English-speaking peers, and if not addressed will continue to contribute to low levels of achievement. It is also clear from the failure to close achievement gaps between both former and continuing EL students and their native English peers over the course of their schooling careers that ongoing assistance is probably needed. It is even more difficult to estimate these costs as few schools provide such ongoing support and there are no known studies that have attempted to assess these costs. Linguistic minorities now comprise more than 20 percent of all students nationwide; in the future this number will grow. It is in the interest of the nation as a whole to invest wisely in these students’ education.

Notes

1. Based on data retrieved January 17, 2014 from the California Department of Education website, Dataquest (http://data1.cde.ca.gov/dataquest/).

2. For example, Gándara and Merino (1993) found schools oft en did not have the resources to reclassify students.

3. Although there are no strict standards to interpret differences in these values, values above 0.8 SD are oft en considered large, values above 0.5 are considered moderate, values above 0.2 SD are considered small, and values below 0.2 are considered inconsequential (see Cohen, 1988, pp. 24–27).

4. Language skills of kindergarten students in the ECLS were assessed by kindergarten teachers, regardless of their primary language, while math achievement was conducted in English and Spanish (National Center for Education Statistics, 2000a). Thus these data can be considered more valid and reliable than assessments of LM students conducted only in English.

5. This section draws heavily from García (2005), pages 92–98.

6. For example, in California, while only 6 percent of English learners received primary language instruction in 2005–06, another 21 percent received primary language support (Data from Dataquest, retrieved February 3, 2007, from http://data1.cde.ca.gov/dataquest/ELP_2_State.asp?RptYear=2005–06&RptType=ELPart2_1x). Funded by a consortium of foundations, Project SOL has collaborated with the Mexican government and the University of Guadalajara to create this online curriculum that has been used in several schools in California.

7. SBAC guidelines, for example, note that “For Mathematics, the Consortium will provide full translations in American Sign Language, stacked translations in Spanish (with the Spanish translation presented directly above the English item), and primary language pop-up glossaries …. However, these have not yet been developed and there are presently no regulations for when they should be provided, nor are there any studies to show how useful the translations are in assisting students. See http://www.smarterbalanced.org/wordpress/wpcontent/uploads/2013/12/SmarterBalanced_Guidelines_FAQ.pdf.

8. The higher weights using the broader population may capture both the effects of poverty and LEP status in the first set of estimates.

9. This raises the question of using bilingual teachers instead of aides and support teachers.

10. In California, for example, 19 percent of students enrolled in grades one through four were poor (meaning their families were at 100 percent of the federal poverty level) according to estimates from the 2005 American Community Survey (see: http://www.census.gov/acs/www/index.html), whereas 49 percent of students enrolled in California public schools participated in the free- and reduced-lunch program (see http://data1.cde.ca.gov/dataquest/APIBase2006/2005Base_SlAptDC.aspx?allcds=000000).

11. In California, poor linguistic minority students continue to lag behind poor, English-only students, which suggest the former group of students may need relatively more resources (Gándara & Rumberger, 2006).
References


*Castañeda v Pickard* 648 F.2d 989 (5th Cir., 1981).


Government Printing office.
Introduction

Rural America is more ubiquitous and its residents more consequential than the caricatures often provided by the popular media. Forty-six million people live in rural communities throughout the United States—communities that encompass 73 percent of the U.S. landmass (Cromartie, 2013). In 2010–11, 57 percent of all public school districts were rural, educating 12 million students, or 24 percent of the school-aged population (National Center for Education Statistics, 2013a).

Numerous studies have addressed the varying needs and challenges of urban students and schools, but educational policymakers and researchers have not shown comparable interest in finance and policy issues specific to rural schools. While the attention on urban centers is warranted, more scholarly attention to rural schools is needed if policymakers are to address meaningfully the needs of these students (see Arnold, 2005; Roscigno, Tomaskovic-Devey, & Crowley, 2006; Schafft & Jackson, 2010; Truscott & Truscott, 2005).

Defining “Rural”

Some researchers define the term “rural” with reference to sociocultural traits including lifestyles, values, and behaviors. Others identify rural communities in terms of occupational or land use patterns, such as agriculture and mining, or draw on the concept of “remoteness” and denote such areas by their population density or access to services and economic centers. The following quote is illustrative of the complexity of rural character: “Rural schools typically share the traits of sparsity of population, low property wealth, small student population, less infrastructure, geographic isolation, and a stronger sense of community” (Dunn, 2003, pp. 102–103).

Federal and state agencies vary in how they define “rural” for policymaking purposes, sometimes differing within a single state (Schwob, 2004). The result of such varied definitions is that it makes setting policy for rural schools complicated, and the comparing of results among studies using different metrics suspect (Brent, Sipple, Killeen, & Wischnowski, 2004). In 2006, the National Center for Education Statistics (NCES) introduced a revised framework for classifying schools. The NCES typology is based on urban-centric criteria and specifies four major locale types: city, suburban, town, and rural. Cities and suburbs are subdivided further as large, midsize, or small. Towns and rural locales are designated fringe, distant, or remote in reference to their proximity to an urban center (National Center for Education Statistics, 2007). The framework provides a defensible and consistent means to identify rural school districts by relying on geographic information systems (GIS) to classify both land areas and the specific location of schools within that territorial region.

Key Challenges Facing Rural Schools

Much research suggests that the sociocultural and ecological conditions that define rural communities and their schools are great strengths. For example, researchers argue that rural schools foster strong, nurturing, and supportive relationships among community members, staff, and students. Some see the remoteness of many
rural schools as providing a focused learning environment without the trappings, influences, and distractions of more populated areas. There is also mounting evidence that supports the value of small schools (see, e.g., Hylden, 2005; Lee & Smith, 1997; Mosteller, 1995; Raywid, 1999). Describing the benefits of small rural schools as the Hobbit Effect, Jimerson (2006) explores why small school size has many benefits (once controlling for SES), including greater academic success, higher graduation rates, more advanced courses, and greater participation in extracurricular activities.

A second rural narrative exists, however, in part to explain the declining trends in rural demographics, economies, and social life (Brown & Schafft, 2011). This opposing narrative describes rural schools as at least partially complicit in the decline of rural communities. Local educators routinely signal to students that the future is elsewhere and the goal of school is to prepare oneself to leave (Corbett, 2007). Rural schools are geared to absorb local, state, and federal investment and expend those resources on the education of children from the community and export them elsewhere. Attention to local assets is not prioritized, and students learn that everything of import is nonlocal (Carr & Kefalas, 2009).

Though measuring the effect of individual, family, community, and school-based factors on student achievement is complicated (see Section II and Section VI, this volume), rural students consistently exhibit lower levels of student achievement and attainment and higher dropout rates than their suburban counterparts (National Center for Education Statistics, 2007). Roscigno, Tomaskovic-Devey, and Crowley (2006), for example, used three waves of the National Education Longitudinal Survey (NELS) and the Common Core Data (CCD) to examine the effect of family resources (e.g., income) and family educational investments (e.g., number of household educational items, cultural capital, and parental involvement) on rural student achievement. Controlling for individual (e.g., gender) and school (e.g., size) attributes, they found that rural families tend to place a lower premium on educational investment than do suburban and urban families with comparable resources. They posit that such differences might be explained by rural histories and cultures (e.g., economic and social stratification). Byun, Meece, and Irvin (2012) also utilized NELS to examine differences between rural and non-rural postsecondary enrollment and completion, demonstrating that rural student attainment was lower on both measures when compared to urban and suburban students. They found that while rural communities provide relatively higher levels of “community social resources,” the low SES of so many rural students suppresses their higher education outcomes.

It is in these contexts that educators encounter significant obstacles when trying to provide their students with improved educational opportunities. In the rest of this chapter we examine four of these challenges and the strategies that rural educators have devised to address them: bounded curriculum, staffing, sparcity, and small enrollments.

Bounded Curriculum

The Challenge

Rural schools are small. Seventy-three percent of remote rural secondary schools enroll fewer than 200 students, and 94 percent enroll fewer than 400 students; 40 percent of distant rural schools enroll fewer than 200 students, and 72 percent enroll fewer than 400 students; and 40 percent of fringe rural secondary schools enroll fewer than 200 students, and 34 percent enroll fewer than 400 students (National Center for Education Statistics, 2013a). The result is that rural secondary schools often do not provide the breadth and depth of course offerings that are available to students in larger systems (e.g., Haller, Monk, Spotted Bear, Griffith, & Moss, 1990). More recently, the NCES (2012a) offered data that demonstrate that a smaller percentage of rural schools offer instruction in various arts subjects (breadth), and within these subjects, fewer courses (depth) than suburban and urban schools. Similarly, fewer rural schools offer Advanced Placement and International Baccalaureate courses (National Center for Education Statistics, 2013b).

Related to the issues of breadth and depth of rural school course offerings is the wave of attention to the implementation of the Common Core State Standards (CCSS). Studies are just underway to examine the implementation issues unique to rural settings (e.g., Walters, 2013), but logic would argue that the same
challenges of quality curriculum implementation would hold and need to be overcome.

Beyond the direct offering of courses, various states and the federal No Child Left Behind Act (NCLB) require supplementary educational services (SES) for struggling students. The provision of these services results in both additional rural advantages and challenges. For example, the sheer number of students requiring additional instruction in multiple subject areas makes it nearly impossible for some urban schools to meet the state requirement during the normal school hours in the space available. Proportionally fewer children require SES services in suburban and rural schools, and rural districts often have fewer space constraints than do suburban and urban schools. While available space and the volume of students in rural areas are not commonly a challenge, the staffing of such services and persistently low performance by some children is a challenge. This is particularly so with the requirement that SES services are to be subcontracted by districts to nondistrict personnel, given the remote location of rural schools (Barley & Wegner, 2010; Killeen & Sipple, 2005; U.S. Department of Education, 2008).

**Strategies**

Distance education refers to a variety of instructional delivery methods that use technology to link students and teachers who are separated geographically. One can place these technologies along a continuum of expanded interactions between teacher and student. At one end, Internet courses may use asynchronous (i.e., not simultaneous) instruction to connect students and teachers (e.g., web-based instructional modules). At the other end, technology connects students and teachers synchronously, enabling real-time communication (e.g., technology-based video conferencing) (Brent, 1999).

Most rural districts have turned to one or more forms of distance education to expand their program offerings. In 2009–10, 59 percent of rural schools offered distance education courses compared with 37 percent of urban schools and 47 percent of suburban schools (National Center for Education Statistics, 2012b). Distance education providers for rural schools included independent vendors, postsecondary institutions, other school districts, and state-sponsored virtual schools. Rural districts used distance education to provide a range of course offerings, such as advanced placement, dual enrollment, credit recovery, career and technical education, as well as other academic courses. The primary impetus for the use of distance education was to “provide courses otherwise not available at school” (National Center for Education Statistics, 2012b, p. 16). Furthermore, 69 percent of rural districts reported that they intended to increase the number of distance education courses they teach in the next three years.

Despite the prevalence of distance education, relatively little is known about the cost-effectiveness of its use generally and in rural schools specifically. With regard to effects, it is clear that distance education enables rural schools to expand their curriculum. Few studies, however, compare student achievement in distance education to face-to-face instruction. Though these studies typically hold that students in distant and traditional instructional settings achieve comparably (Cavanaugh, Gillan, Kromrey, Hess, & Blo-meyer, 2004), methodological limitations make drawing definitive conclusions problematic (U.S. Department of Education, 2010).

In addition, few studies examine the costs that districts incur to provide distance education, regardless of approach (i.e., synchronously or asynchronously). The few studies that do exist reveal that implementing and supporting technology is costly, both in terms of direct expenditures (e.g., capital outlay) and total costs. Brent’s (1999) analysis of interactive television (ITV) in nine rural schools indicated that some districts would have incurred substantially fewer expenditures if they provided classes in traditional settings, albeit in very small classes, rather than through ITV. Brent’s study also revealed a hidden cost of ITV: administrators found scheduling problematic because districts operated on different yearly and daily schedules, including holiday schedules, teacher conference days, and snow days. A survey of 394 rural districts conducted by the National Research Center on Rural Education Support revealed a similar finding: 59 percent of the districts reported that scheduling distance education courses was problematic (Hannum, Irvin, Banks, & Farmer, 2009). This finding may explain, in part, why only 47 percent of their respondents indicated that they were “very satisfied” with distance education.

Though much needs to be learned about the cost-effectiveness of distance education, it is clear that rural schools have become reliant on this mode of instructional delivery (National Center for Education Statistics,
Federal and state funding for distance education explains, in part, this phenomenon (Education Commission of the States, n.d.). It is plausible that the gap in course offerings between rural and other schools will continue to decrease. It is also possible that new and emerging technologies will make distance education a cost-effective alternative to traditional classroom-based instruction.

Staffing

The Challenge

Adequate staffing of schools represents a particularly salient challenge for high-need rural and urban schools. Whether it be recruitment and hiring or retention of high-quality teachers, rural schools face staffing challenges. While discussions of teacher staffing are typically linked to an adequate supply of new teachers (and hence discussions related to the supply line of teacher education), the literature is clear that the challenge is linked to teacher mobility and retention. Nearly one-Fifth of the teachers in the labor force turn over each year with a disproportion of the turnover in high-need urban and rural schools, resulting in disproportionately high teacher turnover costs in these school types (Barnes, Crowe & Schaefer, 2008). Ingersoll and Merrill (2010) provide a comprehensive national analysis using the multiple waves (1987–2008) of the Schools and Staffing survey. They found that 45 percent of all teacher turnover occurred in one-quarter of public schools—with high-need minority urban and rural schools having the highest turnover rates. Moreover, the distribution of teachers today is becoming bi-modal with the junior (zero to three years) and most senior teachers becoming most prevalent. Finally, analysis of teacher moves (from one school to another) shows a "significant annual shuffling” of teachers from poor (rural and urban) to wealthier schools (Ingersoll & Merrill, 2010, p. 19).

The micro labor markets in which school districts operate limit the numbers of prospective teachers (new and experienced) in rural and high-need urban areas. Attempts to model the labor markets for teachers have found them to be more local than was heretofore understood. Researchers found that more than 80 percent of teachers in New York State end up teaching in districts that are located within 40 miles of where they grew up (Boyd, Lankford, Loeb & Wyckoff, 2005). Similar findings were found in Vermont (Killeen, Loeb & Williams, 2013) and in Tennessee, where the vast majority of teachers are employed in a public school near the university where they trained (Tennessee Advisory Commission on Intergovernmental Relations, 2002).

In a study they conducted in Michigan, Maier and Youngs (2009, p. 44) argue that social capital plays a significant role in teacher placement:

Using social networks to hire candidates can also provide benefits to both the employer and the candidate. When a candidate and school are connected through social networks, both have greater access to information about one another—information that cannot be easily gleaned from resumes, cover letters, or possibly even interviews.

This research found, however, that candidates from rural communities were under-represented in university programs. Thus, rural schools may be less likely to be part of the network of student teaching placements and hence out of the social network involving the newest candidates.

Teacher quality is also an important aspect of the teacher labor market. Yet, federal policy defining "highly qualified teachers (HQT)” has been viewed as problematic and even harmful for rural schools (Eppley, 2009; Hammer et al., 2005). In particular, the NCLB requires teachers to be highly qualified in each subject area. As rural teachers are often responsible for teaching multiple subject areas, they and their schools have run afoul of this provision.

The current challenge in policy and practice is to assess teacher quality using student performance data. Driven in large part by the Race-to-the-Top initiative and waivers to NCLB, states are enhancing teacher evaluation by including the use of local or state student assessment data. As of 2012, 20 states and the District of Columbia require student performance data to evaluate teachers (Bornfeund, 2013). In analyzing this process in five states, Bornfeund identifies multiple opportunities and risks with the development of student learning objectives (SLOs) and shared assessments in subjects and grades that are not tested by the state. Among the opportunities are the fostering of school-level collaboration and shared priorities, instructional improvement, and a well-rounded curriculum. The challenges, or as Bornfeund labels “risks,” are that these activities are
human-, time-, and fiscal-resource intensive and schools have limited expertise. While Bornfeund did not single out rural districts, these challenges are, no doubt, straining rural school staffs and budgets.

Distance to key services, such as libraries, universities, health services and hospitals, and even shopping opportunities, represent additional obstacles for the staffing of rural schools. More than just an inconvenience, lack of access to these services signifies a substantial cost to individuals and districts. Lack of access to a larger community or university library may impede teacher learning. A teacher who must drive for an hour in order to take his or her child to the hospital will miss a greater proportion of a workday than someone who has a five-minute drive to an appointment. The district then faces an additional cost of finding a qualified substitute teacher for the time the regular teacher is out of the building. Finally, the students bear the brunt of the cost of a less productive day of school with a substitute teacher.

Given what is known about the rural teacher labor markets, questions of incentives arise. What incentives and other strategies are known to alter the recruitment, hiring, retention and quality of educators in rural areas?

**Strategies**

Approaches to reducing social isolation are straightforward, though the cost-effectiveness of such strategies in rural areas is unclear. Strategies include the use of technology, financial incentives, and regional partnerships. As noted above, technology is now in place in many rural areas to offer experiences, coursework, and enrichment to students and their teachers.

The research on teacher recruitment strategies shows that policymakers use signing bonuses and other financial incentives to staff hard-to-fill positions (Greenlee & Brown, 2009). In a review of strategies used by different states, Clewell et al. (2000) summarized policies that have helped local school districts’ efforts to recruit more and better teachers. These included loan forgiveness, salary increases, signing bonus, and allowing the retired teachers to draw pensions if they return to the classroom. Spradlin and Prendergast (2006) documented the number of states that offered various incentives to assist in teacher recruitment, including scholarship programs (32 states), loan forgiveness (39 states), alternate path to certification programs (47 states), recruitment bonus/housing allowance (14 states), and grow-your-own initiatives (11 states). Among these programs, few targeted specifically rural schools.

Teacher and administrator salaries in rural areas are lower than in suburban and urban districts, though differential compensation is necessary to attract teachers into remote regions located away from the facilities and services of urban life. In a national study of the costs and benefits of altering teacher salaries, Chambers (1995) concluded that salary incentives for rural schools are necessary to overcome what Jimerson (2003) termed the "competitive disadvantage" of teacher compensation in rural areas. Chambers’ study calculated that the "average student" attending a rural school would have access to the "average teacher" at a cost savings of 8 percent when compared to the non-rural average. Another student attending a "remotely located" district would have access to that same quality teacher at about the average (same) cost of a non-rural school. Hence, to offset the reduced labor market in remote areas, actual labor costs (salaries) should match that of the less-rural school districts.

Regional educational agencies also offset the challenges of providing adequate staffing through the provision of staff development opportunities. Regional labs and intermediate educational organizations serving multiple component districts (e.g., Intermediate School Districts in Michigan and Boards of Cooperative Educational Services in New York State) provide valuable staff development to rural schools (Sipple, 2004; Sipple & Killeen, 2004; Sipple, Killeen, & Monk, 2004). Policies that require minimum hours of professional development, provide funding, and establish programs (such as regional service centers) all contribute in positive ways to the promotion and delivery of higher-quality professional development.

Hammer et al. (2005) offer a promising set of strategies, including some rural-specific themes, to improve the recruitment and retention of teachers in rural schools. Overall, the strategies for effective recruitment and retention must be strategic, targeted toward the hard-to-staff grades and subject areas, sustained year after year, and grounded in the local community. Those strategies found to be effective in attracting and retaining teachers include basing recruitment on data, building partnerships, offering targeted incentives, regular evaluation, using building-level staff in hiring, formal induction programs, and involving the community in the
welcoming of new hires.

The use of emerging technology can also reduce the especially heavy recruiting and hiring costs to small rural districts. Districts can purchase access to web-based services to facilitate efficient advertising, recruiting, communication, and reviewing potential candidates for teaching positions. While large districts offer the economies of scale to hire human resource professionals, very small districts rely on administrators and teachers to do the advertising, recruiting, and hiring. The cost of advertising in a regional newspaper is the same for a district of 300 as it is for a district of 10,000. One example of such a service is SchoolSpring (schoolspring.com), a private service that posts positions from across the country for a fee and then provides an archiving service for application materials.

Using national data from the Schools and Staffing Survey, Beesley, Atwill, Blair, and Barley (2008) analyzed strategies for recruitment and retention of rural secondary teachers. Rural school principals reported greater recruiting and hiring challenges than did non-rural principals in the central United States. The most successful practices were grow-your-own teacher recruitment, the use of federal dollars to support aides and other noncertified teachers to earn certification, and some targeted financial incentives to enhance recruitment and retention.

The use of alternative pathways to teaching in rural areas warrants increased attention. One alternative paThis Teach for America (TFA) which places their corps members in rural areas. An early study conducted by Mathematica found that TFA teachers are virtually indistinguishable (in terms of student learning gains) from their traditionally certified colleagues in high-need rural and urban schools. One distinguishing difference was a modest though positive effect of TFA mathematics teachers when compared to non-TFA mathematics teachers (Glazerman, Mayer & Decker, 2006). A subsequent review of an array of studies on TFA supported these findings (Heilig & Jez, 2010).

An updated study by Mathematica compared student mathematics performance when taught by TFA, Teaching Fellows Programs, and other more traditionally trained teachers. Students with TFA teachers performed statistically better (0.07 SD) than the comparison teachers. Students with Teaching Fellow teachers performed no different than other teachers, but did outperform students taught by teachers from “less selective” teacher pathways (Clark et al., 2013). Two additional studies found that students of TFA teachers outperformed students with non-TFA teachers in mathematics (Antecol, Eren, & Ozbeklik, 2013) and science (Xu, Hannaway, & Taylor, 2011).

Variation within preparation path, however, is substantial and may be as important to understand as the mean differences between groups. In New York, Boyd et al. (2005) concluded that the variation in effectiveness within pathways is far greater than the average differences between pathways. Antecol, Eren, and Ozbeklik (2013) also emphasize the heterogeneity within preparation pathways, particularly TFA.

While the literature describes a wide array of strategies to enhance the staffing of rural schools, the number and quality of relevant policy evaluation studies is limited. Hence, the long-term effect of using these strategies and incentives to increase the size of an applicant pool, to recruit a teacher once an offer is made, and to retain teachers in remote locations is not well understood.

\section*{Sparsity}

\subsection*{The Challenge}

The cost of school transportation has increased since the 1930s, punctuated by periods of substantial growth. From 2001–02 to 2010–11, the average cost per student transported increased by 19 percent in constant dollars (National Center for Education Statistics, 2012c). Such a large increase in a noninstructional function is particularly burdensome for rural schools (Strange, Johnson, Showalter, & Klein, 2012). Because of their larger catchment areas, rural schools devote 40 percent more of their operating budget to transportation than do urban schools and 18 percent more than suburban schools (National Center for Education Statistics, 2012c; see also Killeen & Sipple, 2000).

In addition to comparatively higher transportation expenditures, rural districts’ larger catchment areas place
a hidden cost on students. Howley, Howley, and Shamblen (2001) reported that rural students often travel an hour or more to and from school. Such lengthy bus rides pose challenges to student well-being and educational programming. A study of the effects of long bus rides in rural Ontario revealed reduced student sleep, recreation time, academic performance, and extracurricular participation (Howley et al., 2001). Howley et al. (2001) also argue that long bus rides may well erode rural communities’ social fabric. Indeed, it is plausible to argue that as the distance from school to household increases, parents as well as students may be less likely to engage in school-related activities. Interestingly, the NCES (2007) reports that a greater percentage of rural parents (74 percent) attended a school event in the past year, when compared to urban (65 percent) and suburban (71 percent) parents. While there is a paucity of research specific to the nature and effect of rural school parental involvement, the literature is robust as to the value of parental involvement more generally.

**Strategies**

Transportation costs impacted by sparsely populated areas can be offset by state aid. In a comprehensive survey of state aid programs, Verstegen (2011, p. 22) identified six general approaches to pupil transportation funding:

1. a separate calculation, or part of a block grant in the general state aid formula (9 states);
2. density formulae based on bus-route miles, pupils per bus-route mile, or square miles in the district (9 states);
3. cost reimbursement formulae with a fiscal-equalization feature to adjust disbursement of funds to school districts (3 states);
4. cost-reimbursement formulae that pay full cost to school districts (5 states);
5. cost reimbursements that only reimburse the district for approved costs (17 states);
6. programs that pay a uniform amount for each transported pupil (5 states); and
7. no state funding for transportation (2 states).

State-specific transportation funding can vary considerably within a given approach. Nevertheless, some approaches should ameliorate the fiscal burden of large catchment areas more than others. Full-cost reimbursement approaches relieve rural districts, as well as other locales, of all transportation costs. Density-formulae approaches strive to address directly the additional transportation costs that derive from sparsity, though reimbursement levels may fall short of the full cost of transportation. Several per-pupil formulae, an approach that would appear, on its face, to disadvantage rural districts, allocate greater amounts per pupil based on the distance in miles between a student’s residence and school (Verstegen & Jordan, 2008). Like density formulae, graduated-per-pupil formulae are unlikely to mitigate fully the additional transportation costs borne by geographically large but sparsely populated districts.

One cannot generalize defensibly about the effects of approved cost-reimbursement approaches on rural schools, as what is considered an allowable cost varies among the states, as do levels of reimbursement (e.g., 60 percent or 90 percent). Similarly, the nature of non-categorical aid-funding approaches confound efforts to determine how state and local revenues combine to affect transportation spending decisions.

**Small Enrollment**

**The Challenge**

Debate over the costs and benefits of small schools has raged for a century. Improving technology and growing support for the small-school movement (Hylden, 2005) has tempered, somewhat, concerns about the educational opportunities available to rural students. However, little has changed over the past 100 years to quell concerns about the perceived diseconomies of scale (i.e., inefficiency) inherent in small districts (Kannapel & DeYoung, 1999).

**Strategies**

Many state-level policymakers subscribe to the belief that consolidation (i.e., the merging of two or more districts into a single, larger district) is a means to improve the cost-effectiveness of rural schools, though there
is little evidence that such actions improve Efficiency. For example, Arkansas’ 2004 Public Education Reorganization Act (Act 60) required that any school district with an enrollment of fewer than 350 students either (1) consolidate with one or more other district(s) to create a new district that would meet the minimum size requirements, or (2) be annexed into an existing district meeting those requirements (Johnson, 2006). Other states encourage consolidation by providing incentives through operating or building-aid programs (Lawrence et al., 2002; Sielke et al., 2001).

Scholars have examined whether the resulting consolidated districts are, or are likely to be, more efficient. Cost-effectiveness analysis techniques (demanding that the effects and costs of policy actions be considered) suggest that the efficacy of these strategies is at best ambiguous. Proxies for “educational effectiveness” have included measures of student achievement, student social development (e.g., extracurricular participation), graduation rates, school climate (e.g., disciplinary actions and student and teacher attitudes), curriculum comprehensiveness, and community responsiveness (see, e.g., Fowler & Walberg, 1991; Haller, 1992; Haller & Monk, 1988; Howley, 1989; Howley & Howley, 2004; Nathan & Fehey, 2001; Verstegen, 1990). To date, researchers offer little convincing evidence that students of small rural schools have a lesser educational experience than those in larger schools, with one important exception—curriculum comprehensiveness (e.g., Anderson & Chang, 2011; Graham, 2009). The Carsey Institute study finds that, when lacking opportunity to advance to higher level academic courses, rural students in small schools have lower scores on assessment tests resulting in fewer qualified students accessing science, technology, engineering, and math (STEM) job pipelines (Graham, 2009). Others show similar limitations of mathematics course-taking and preparation in rural versus non-rural schools (Anderson & Chang, 2011; Redding & Walberg, 2012).

Scholars have also considered how district consolidation affects rural communities more broadly. Some studies posit that the loss of a rural school adversely affects the community’s economy by eliminating a sizable employer and thereby fostering a decline in retail sales, property values, and tax revenues (Li, Francis & Sipple, 2013; Lyson, 2002; Sederberg, 1987; Sell & Leistritz, 1996; Sipple, Anantsuksomsri, Tontisirin, & Francis, 2013). Hu and Yinger (2008) demonstrate that how district consolidation affects housing values is a function of the size of the districts that consolidate, state aid provided as an incentive to consolidate, and neighborhood wealth. Rural schools are often the center of community social life, offering a reason and site for recreational, cultural, and civic events (e.g., sports, plays, dances, and political meetings) (Lyson, 2002; Peshkin, 1978). In replicating Lyson’s work, researchers found that the presence of a school building in small rural communities had a stabilizing effect from 1990 to 2010 on local property values and incomes, when compared with those communities without a school (Li, Francis, & Sipple, 2013; Sipple et al., 2013). In measuring the actual distance from village center to the nearest school, it is apparent that the farther away the school the greater loss of property values and income (Sipple et al., 2013).

There is a dearth of work on how consolidation affects students, teachers, and administrators who transition from independent to combined districts. Nitta, Holley, and Wrobel (2010) conducted four case studies of consolidated districts in Arkansas. Their findings suggest that students adapted better socially than did teachers. Though most teachers reported that they experienced improved working conditions and professional development, they also noted difficulty in developing relationships with colleagues and students.

In principle, larger schools should be able to offer a given instructional program at a lower expense per pupil than can smaller schools. Scholars have often nested discussions of the costs of consolidation within broader analyses of size economies in education (Duncombe, Minor, & Ruggiero, 1995; Tholkes & Sederberg, 1990; Zimmer, DeBoer, & Hirth, 2009). Duncombe and Yinger (2010) report these efforts typically indicate that administrative and instructional savings may result from “moving from a very small district ... to a district with 2,000 to 4,000 students” (p. 11). They note, however, that these efforts do not examine expenditures differences before and after consolidation, but rather the relationship between district enrollment and per-pupil expenditures, typically disaggregated by broad functional account areas (e.g., administration, instruction, and transportation). Duncombe and Yinger (2007) addressed this limitation by directly studying changes in expenditures per pupil for 12 pairs of rural districts that consolidated in New York State between 1985 and 1997. Their analyses reveal decreases in instructional and administrative expenditures per pupil, though the level of savings decreases as the size of the consolidating districts increases.

Many policymakers also argue that educational service agencies (ESA) offer a means to overcome diseconomies of scale in small rural schools. Known variously as Boards of Cooperative Educational Services
(New York), Intermediate School Districts (Michigan), and Educational Service Centers (Texas), ESAs are organizational units positioned between districts and the state. Though these units vary greatly regarding their governance and organizational structure, all are expected to provide services to component districts more cost-effectively than if the districts provided these services on their own (see, e.g., Stephens & Harmon, 1996; see also Fletcher & Cole, 1992; Galvin, 1995; Helge, 1984).

One or more of the following features typically characterize ESA services (Stanley, 1995; Stephens & Harmon, 1996): costly services that necessitate a concentration of students (e.g., programs for disabled students); services that warrant staff expertise that is in limited supply (e.g., curriculum development); and services that require expensive equipment or facilities (e.g., vocational education). Despite its raison d’être, the ESA’s ability to improve rural school efficiency has been investigated in only a few studies. Nevertheless, three tangible lines of evidence suggest that ESAs might be cost-effective in assisting rural school districts. First, researchers have found that component districts are satisfied with the quality of services provided by their ESA (see, e.g., Hayden & Fielder, 1999; McKinney & Gauntt, 2000; Thomas, 1997). Second, several studies have compared the costs of locally provided versus ESA-provided services, typically favoring districts’ use of ESAs (see, e.g., Campbell, 2001; Wallerstein, 1997). Thirdly, a few studies have documented that administrators believe that ESAs provide services more cost-effectively (i.e., efficiently) than do individual districts (Stanley, 1995; Thomas, 1997; ECM, Inc., 1997).

Most states provide aid to districts when they use ESAs. For example, New York State provides districts with expense-based aid (equalized by relative property wealth) to purchase ESA services (University of the State of New York, 2013). This policy enables the district to pass along a portion of its local cost to the state. Thus, although it may be less costly for a district to contract for services because it is aid-able, it does not follow that the total cost to the broader system is always less than if it provided these services on its own. We are aware of no study that accounts for the role of state aid when studying the cost-effectiveness of ESAs.

**Conclusion**

In reviewing the unique challenges facing rural schools and the effective policies and strategies to assist in overcoming the challenges, we continue to acknowledge the lack of research on rural schools. While much research has targeted the challenges and policies related to quality curriculum and now the Common Core State Standards, teacher recruitment and now teacher evaluation linked to student test scores, school size, and scale economies generally, the explicit study of these issues in rural settings is rare. Without more and better research that is targeted on rural populations and takes into account the remote location, sparsity, and size of rural schools and communities, there is a danger in perpetuating the stereotypes, misunderstandings, and misdirected policy initiatives surrounding rural schools.

It will be imperative for researchers and policymakers to pay specific attention to the implementation of the Common Core and the emphasis on enhanced teacher evaluation in rural settings. Each of these policy reforms requires additional resources at the same time that districts have pared down budgets to survive the recession of 2008–10. Each of the four challenges we have discussed, and the strategies aimed to overcome the challenges, generates an array of future research questions. For instance, what are the long-term costs and benefits of the use of technology to reduce social and academic isolation in remote areas? Such work has implications for distance learning, the reduction of teacher and student isolation, staff turnover, and transportation costs. To what degree does distance traveled from home to school impact student achievement, participation in extracurricular activities, parental involvement in schools, and local community vitality indicators? In light of the rapid technological advances in personal and group communication, how does the new technological environment impact the scale economies in schools?

Another major change taking place in some rural communities is changing demographics (Brown & Schafft, 2011). These changes are uneven across rural communities, but in those communities with greater numbers of in-migrants, what are the implications for students and communities?

The findings of such research may document the disadvantage of rural schools and press for further consolidation. Conversely, as broadband access to rural communities grows, or population increases, it may reduce the isolation and turnover of teachers and students and reverse trends in school closures. These are just
a few of the many research and evaluation studies necessary to fully understand the challenges confronting rural schools and their communities and to assist policymakers in their quest for more equitable and adequate school resource policies.

References


The Organizational and Fiscal Implications of Transient Student Populations

KIERAN M. KILLEEN AND KAI A. SCHAFFT

Student Transiency: Defining the Issues

Student transiency is a particularly complex issue in education. This chapter offers an introduction to historical research about transiency, and then discusses how transiency creates distinct student, school, and related organizational impacts. Student transiency refers to the repeated nonpromotional and unscheduled movement of students from one school or school district to another. Authors typically use the term "student mobility" to refer to such school changes, but this chapter uses "transiency" to widen attention to the multidimensional nature of school changes and stress the complexity of making effective policy choices in response to such student movements. In short, highly transient students tend to come from low-income families, migrant or limited-English-proficiency backgrounds, and/or single-parent families (Ashby, 2010; Nevárez-La Torre, 2011; Rumberger, 2011; Rumberger, Larson, Ream, & Palardy, 1999). The lives of transient children are socially and academically disrupted through these frequent and unpredictable school and residence changes (Grigg, 2012; Herbers, 2013). And these movements create direct and indirect challenges or costs in education.

Census data show that about 15 percent of the U.S. population changes residence each year, and between 2005 and 2010 over 35 percent experienced at least one residential change (Ihrke & Faber, 2012; Ihrke, Faber, & Koerber, 2011). Americans frequently think of residential mobility as a voluntary and opportunity-related behavior in which people choose to move in order to take advantage of social and economic opportunities at migration destinations, and this generality is accurate with respect to the movement of households with relatively high and stable socioeconomic status. However, demographers have long noted that households at the wealthiest and poorest ends of the spectrum are most likely to be residentially mobile (Phinney, 2013; Ritchey, 1976). Unlike wealthier households, low-income families often move for largely unplanned and unpredictable reasons, including household crises such as family breakup, inability to pay rent, or movement away from unsafe, unaffordable, or otherwise unacceptable living conditions (Schafft, 2005 and 2006).

Despite significantly negative consequences for both children and schools, student transiency is underrecognized within the research and policy arenas (Hartman & Franke, 2003). This is the case for several reasons. First, transient students are an institutionally untargeted population. For example, federal aid for education flows towards schools serving children of migrant workers and homeless students. Although migrant and homeless students may often experience frequent residential mobility, many students from poor, mobile households do not fit the sociodemographic profiles of either migrant or homeless students (Hartman, 2002; Schafft, 2006). In fact, poverty-related student mobility is often highly localized and may have no connection to migrant labor and/or seasonal agricultural economies (Schafft, 2005 and 2006). Similarly, while low-income transient students may often live in temporary and/or inadequate housing, they may not technically be homeless.

Additionally, although schools may experience significant churning of student populations, the numbers of entering and exiting students are often roughly equal and thus in many instances have only negligible effects on net enrollment (Schafft, 2005). This further contributes to the “invisibility” of student movement. As a superintendent in a rural New York district remarked on the incidence of transiency within his district (Schafft, 2005, p. 10):

Most of the community does not recognize (student movement) as an issue. But it creates a huge problem … the aid is frozen by the state. To
pay for the needs of these kids we will have to go to the local taxpayer. It’s a hard sell to the community at large that we have this unknown group that requires some substantial resources that don’t even exist to most people here but nonetheless are very real to us.

Despite the relative invisibility of these students, however, a number of studies have identified the scope and magnitude of student transiency, especially within economically disadvantaged school districts and among economically disadvantaged students. A study that followed a nationally representative kindergarten student cohort from 1998 through 2007 found that 18 percent of students had changed schools three times by the eighth grade and that 13 percent had changed schools four or more times (Ashby, 2010). The students in the study who had changed school four or more times differed significantly from their peers who had changed schools two or fewer times. More-mobile students tended to be disproportionately low-income, African American, and from single-parent and non-home-owning households. An Ohio study found that one-year district level mobility rates were often 20 percent or more and at the school building level, rates of 40 percent or higher were not uncommon, especially in low-income urban schools (Garber, Dowell, Keithley, Ryan, & Partin, 2012). These findings echo those of an earlier California study showing that one out of five high schools was found to have annual mobility rates greater than 30 percent, and one in 10 had mobility rates greater than 40 percent (Rumberger et al., 1999).

Student transiency has serious implications for students and schools and deserves further attention within educational, administrative, and public policy arenas. It disrupts the social and academic lives of students, reduces academic achievement, and leads to a higher risk of dropping out. Student transiency also creates distinct fiscal, administrative, and record-keeping challenges for schools. In an era of high-stakes testing, schools face additional threats of sanctions resulting from the test scores of low-achieving, high-need students who may well have spent only limited time receiving academic preparation from the school or district being held accountable for their learning (Schafft, Killeen, & Morrissey, 2010).

This chapter is organized in two main parts. First we review the research evidence regarding the direct effects of transiency on students’ social and academic outcomes. We then discuss specific resource and organizational implications of transiency. This latter part stresses the complexity of transiency as a particular policy problem in educational finance and policy studies.

The Social and Academic Impacts of Transiency

Research consistently points to profoundly negative academic and social outcomes associated with student transiency. In 2010, the National Academies of Science convened a group of experts to reexamine the extant literature on the relationship between student mobility and achievement (Beatty, 2010). The report restated the consistent negative correlation between transiency and a variety of academic outcomes for youth that appears widely in the research literature (see also Evans, Yoo, & Sipple, 2010; Gasper, 2012; Grigg, 2012; Herbers, 2013; Ingersoll, Scamman, & Eckerling, 1999; Lamote, 2013; Lash & Kirkpatrick, 1990; Mehana & Reynolds, 2004). Authors also consistently associate transiency with adolescent behavioral problems, particularly as a function of disrupted social networks and social relationships (Osher, Morrison, & Bailey, 2003; Tucker, Marx, & Long, 1998). Social and behavioral problems include attention deficits, increased risk of school suspension, delinquent rule-breaking behavior, and social difficulties (Engec, 2006; Ersing, Sutphen, & Loeffler, 2009; Lamote, 2013; South, Haynie, & Bose, 2007). The harmful effects of mobility are more pronounced among older adolescents and among girls and appear to persist for several years (South & Haynie, 2004). Research has also linked mobility with early instances of sexual intercourse (Haynie, South, & Bose, 2006) and victimization (Carson, Esbensen, & Taylor, 2013).

Typical research about student mobility tends to relate academic outcomes with both mobile and nonmobile students. For example, using data from a nationally stratified sample of elementary school children in the third grade, the U.S. General Accounting Office (1994) reported that about 41 percent of highly mobile students scored below grade level in reading, and about 31 percent scored below grade level in math. In comparison, only 26 percent of stable students (those who attended only one school since kindergarten) tested below grade level in reading, and about 16 percent tested below grade level in math. The study also found that highly mobile third graders were far more likely than nonmobile students to repeat a grade.

Frequent school changes also impact nonmobile peers. As transient students are integrated and reintegrated
into classroom instruction, attention and resources are diverted away from resident students (Lash & Kirkpatrick, 1990). It not just that changing schools can create weaker academic outcomes for children, the movement of children in and out of classrooms creates challenging teaching environments. Highly mobile schoolchildren tend to exhibit increased externalizing behavior, or exhibit more frequent behavioral problems, often with other students (Malmgren & Gagnon, 2005). High mobility rates at the school level are associated with increased risk of dropping out among nonmobile students as well as mobile students (South, Haynie, & Bose, 2007; cf. Chen, 2008).

On a more conceptual level, the academic impact of changing schools may be thought of as a particular form of invisible absenteeism. Unlike a continuously enrolled child who may temporarily withdraw from school for illness, appointments, or pressing family engagements, transient students often experience periods of nonenrollment during transition from one school to another. This itself is a form of truancy, truancy that may not actually monitored by the schools. In a study of student mobility in the Pittsburgh (PA) City School District, researchers standardized the impact of mobility by the equivalent academic effect of classroom absence. This study found that a change in schools in year one had the academic equivalency to being absent 32 days in year two, and 14 days in year three (Dunn, Kadane, & Garrow, 2003; cf. Grigg, 2012).

Despite this consistency in the literature base, there exists a small set of studies that question the direct and independent effects of transiency among other factors that tend to depress social and academic outcomes. Studies that control extensively for measures of student background and family context tend to observe diminished effects of mobility on child outcomes (Hanushek, Kain, & Rivkin, 2004; Pribesh & Downey, 1999; Swanson & Schneider, 1999). Specifically, investigations using longitudinal data observe that negative childhood academic or developmental outcomes are associated with early childhood experiences prior to school enrollment, such as participation in an early intervention program and family poverty, even when controls are added for student mobility in later grades (Temple & Reynolds, 1999). Work by Dong et al. (2005) extends this finding. Using adolescent health outcomes, these authors suggest that mobility acts as a proxy for very compelling, adverse childhood experiences such as general maltreatment, sexual abuse, or family dysfunction. Each of these precipitating factors of adverse early childhood experiences likely contribute to poor outcomes during adolescence, outcomes that appear to interact and at the very least are worsened by student mobility. Overall, this research suggests how mobility disrupts social and academic stability, and it underscores the need to better understand its causes, consequences, and the most appropriate responses for policy and practice.

Students Most Likely to be Mobile

Student transiency is highly associated with household-level economic distress and tends to occur with greater intensity in economically disadvantaged communities and neighborhoods. However, other student subpopulations are also highly mobile for a variety of reasons, sometimes related to economic distress, and sometimes not. These subpopulations include migrant students, homeless students, and children from military families. We discuss each in turn, concluding with a discussion of chronically mobile students from poverty backgrounds who have not been identified as migrant or homeless students.

Students from Migrant and Immigrant Backgrounds

American society is becoming increasingly ethnically diverse, in large part due to new waves of immigrant populations (Lichter & Johnson, 2006; Lichter, Parisi, Taquino, & Grice, 2010; Kandel & Cromartie, 2004). Among immigrant populations, families employed within agriculture or related industries are those most likely to be mobile. Data from the Title I Migrant Education Program for the 2001–02 academic year indicate that the migrant student population was nearly 900,000, a growth of 11 percent from three years earlier, with just over half of that population in California, Texas, and Florida. Ninety percent were Hispanic and one-third had limited English proficiency. Over one-half received some type or combination of social work, advocacy, or outreach services (U.S. Department of Education, 2006).
Seasonal farm laborers that work field crops, nurseries, and livestock tend to earn well under the federal poverty line; in terms of 2011 median hourly wages, such laborers earn $8.99 per hour, which is 60 percent lower than construction laborers by comparison (USDA-ERS, 2014). For this reason migrant students are disproportionately more likely to live in poverty than the nonmigrant student population. Immigrant and migrant students must therefore confront challenges of poverty and academic instability in addition to linguistic and cultural barriers to educational achievement (Green, 2003; Hanna, 2003). These challenges are reflected in significantly reduced levels of academic achievement among migrant students. In the late 1990s only 40 percent of migrant students entered ninth grade in comparison to 96 percent of nonmigrant students. Similarly, only 11 percent of migrant students entered 12th grade as compared with 80 percent among nonmigrant students (Olsen & Jaramillo, 1999). These outcomes are apparent despite intensive federal interventions in the form of Migrant Education Programs (MEP), in which children aged three to 21 are served with in-school and out-of-school education services by caseworkers (U.S. Department of Education, 2006).

A more recent trend is for migrant families to relocate to areas that offer work in physically demanding and low-paying meat-packing or food-processing plants, particularly in the rural South and Midwest (Dalla & Christenson, 2005; Saenz & Torres, 2003). These industries have increasingly established themselves in declining rural communities that are eager to attract new economic activity. With little to no presence of organized labor, these plants can offer low wages and target foreign and migrant (and often undocumented) workers to fill labor needs. The increased population diversity can create tension, particularly in communities and neighborhoods that historically have been ethnically homogeneous (Dalla & Christianson, 2005) and ill-equipped to educate immigrant children (Saenz & Torres, 2003). Rapid population growth of immigrant and migrant families can strain community solidarity, as well as a community’s capacity to provide housing, education, and social services. This is especially the case as the fiscal burdens for these sectors in the past several decades have shifted increasingly to state and local revenue sources (Dalla & Christenson, 2005).

Nonetheless, structures are in place to assist migrant students and their families. By law, school administrators must identify migrant children eligible for receiving services, and school funds are available to support recruitment efforts to actively enroll migrant students (Green, 2003). The Migrant Education Program, part of the Title I program, is a federally funded effort that offers supplementary instruction and supportive services, and is structured to address the interrelationships between poverty, mobility, and academic underachievement (Branz-Spall, Rosenthal, & Wright, 2003). While unable to meet all the needs of migrant children, the Migrant Education Program has proven an important resource for schools and districts in developing flexible programming and in facilitating records transfer for highly mobile migrant students. However, migrant students face overwhelming social, academic, and economic odds. Academic achievement gaps continue to starkly illustrate how much more progress still must be made in the effort to provide adequate and appropriate education to migrant students.

**Homeless Students**

The National Center on Family Homelessness (2011) estimates that 1.6 million children in the United States are likely to experience homelessness on an annual basis. Popular stereotypes of what constitutes homelessness tend to be urban-oriented, and conjure up images of street people and those staying in inner-city homeless shelters. The federal definition of homelessness among youth, however, covers a much wider range of housing-insecure situations. These situations include children and youth who lack a regular or adequate nighttime residence; those sharing the housing of others (or “doubling up”) because of housing loss or economic hardship; and children in substandard housing or living in motels, trailer parks, or campgrounds because of the lack of other, alternative housing arrangements. While this is a broad definition of homelessness, it also leaves much to interpretation, and identification of homeless children requires a more detailed knowledge of a child’s living circumstances than may be available to school personnel.

The federal McKinney-Vento Homeless Assistance Act of 2002 mandates that every school district designate a local “liaison” for homeless children and youth, often a guidance counselor, social worker, or school psychologist, who is responsible for identifying and advocating for students living in homeless circumstances. In the broadest sense, the liaison’s responsibility is to ensure that homeless children and youth have the same opportunities for academic success as their nonhomeless peers. Local liaisons facilitate school enrollment,
obtain student records, and ensure that homeless students have appropriate access to school and support services. The McKinney-Vento Act also requires that districts provide transportation services to reduce the academic transiency of Residentially mobile homeless children. If a child becomes homeless and relocates outside the school district, the student is entitled to remain in the district of origin for the duration of the homelessness, in accordance with state education department plans and policies determining how districts will apportion the provision of transportation to homeless youth.

Some researchers argue that more information is needed within schools and communities about the services to which homeless children and families are entitled and about who is entitled and under which circumstances (Julianelle & Foscarinis, 2003; Schafft, 2006). Homeless parents are rarely aware of rights guaranteed to their children, and districts are often in noncompliance with these regulations because they are not fully aware of either the circumstances of the child or the specifics of the law. For example, 13 New York school districts lost a 2004 federal class action lawsuit when they were found to have obstructed homeless students from enrolling in school by not providing adequate transportation. This denial of service resulted in a lack of access to school for up to months at a time for some of the plaintiffs. Thus, school districts must develop protocols to ensure that students who meet the criteria as homeless are accorded the rights guaranteed them and are provided with the opportunity to remain within the school district despite uncertain housing circumstances.

**Mobile Students From Military Families**

Regardless of the type of school in which they are enrolled—public, private, or Department of Defense schools, children from military families are highly mobile. Schools within the Department of Defense Education Activity (DoDEA) system, for example, typically experience 35 percent or greater student turnover annually (Smrekar, Guthrie, Owens, & Sims, 2001; Smrekar & Owens, 2003). In addition, children from military families often have sociodemographic backgrounds frequently associated with lower academic achievement. Eighty percent of all children in Department of Defense schools have a parent who is enlisted, and most enlisted personnel have no greater than a high school diploma and incomes at or near the poverty line (Smrekar & Owens, 2003). Additionally, 40 percent of the students are from minority backgrounds (Military Family Resource Center, 2001), and as many as 50 percent of students in Department of Defense schools qualify for free or reduced-price lunch (Smrekar, Guthrie, Owens, & Sims, 2001).

Despite elevated mobility and other factors that would generally be associated with decreased academic achievement, children from military families, and especially those attending DoDEA schools, typically score at or above national averages when taking into account the level of parental educational attainment (Popp, Stronge, & Hindman, 2003; Smrekar & Owens, 2003). Several factors likely account for this performance. About 90 percent of children in DoDEA schools come from two-parent families in comparison to about 70 percent of children in public schools nationwide. In addition, these children, while mobile, come from families with a steady source of income, and housing is provided if families live on military bases. Therefore, children from military families do not experience the chronic physical and psychological stress that accompanies prolonged economic and housing insecurity. Additionally, military communities are tight-knit, often with family support mechanisms to ease the stress of transitions associated with residential relocations and deployments (Popp et al., 2003; Smrekar & Owens, 2003). Last, there is no social stigma attached to family and student mobility associated with military family relocation as there is with poverty-associated mobility (Popp et al., 2003).

DoDEA schools, in particular, are structured to handle the social and academic needs of mobile students and accommodate the associated administrative and fiscal demands. Measures include full-time registration and records clerks as well as institutionalized orientation procedures for new students. Family Readiness groups and Family Care plans, coordinated through school and military units, provide care and support systems for children when parents are deployed. There is also a high level of expected parental involvement in DoDEA schools. Military personnel are granted leave time to take part in parent-teacher conferences, as well as to volunteer in the school each month if they choose to do so (Smrekar & Owens, 2003). In sum, while mobile children from military families experience the psychological and social stress of frequent moves, support systems are in place socially and academically to help children adjust, and movement is neither directly caused by, nor a contributing factor to, family disorganization.
Transiency and Poverty-Related Household Mobility

Unlike students from military or migrant farm-working families, other students are mobile due to poverty-related reasons and the inability to secure adequate, affordable housing. These conditions have been worsened by declining rates of homeownership among low-income families, rising rent burdens, limited access to housing assistance, and short-term tenancy in overcrowded and often unsafe housing (Crowley, 2003; Fitchen, 1995; Schafft, 2006). More recently, Allen (2013) examined relationships between childhood mobility and housing foreclosure among households in the Minneapolis City School System and found that one-third of children living through foreclosure also experienced mobility out of the school system. In essence, housing insecure families do not move in response to economic or other opportunities. Rather, these families move when they find themselves without housing or having to vacate current housing because of safety issues and/or affordability. The number of transient students mobile for poverty-related reasons is difficult to estimate. However, students living in unstable housing arguably constitute America’s largest group of chronically mobile students.

Because they are an institutionally untargeted group, transient students face particular disadvantages. Unlike homeless and migrant students, or students from military backgrounds, there are neither programs nor mandates to meet the particular academic and social needs of the transient student who, while not technically “homeless,” may frequently change residence and schools for reasons of poverty and housing insecurity. Regardless, transient, low-income students are at high risk of social and academic problems. As a guidance counselor from an Upstate New York school district explained, “No one owns these kids. They have no political or economic power. The chances of reform happening (for them) are certainly less than they might be for other groups” (Schafft, 2005, p. 11).

Schools and Settings Most Likely to Experience Student Transiency

Arguably the single most significant predictor of high student mobility is poverty—that the household level and at the school and community level. Seventeen percent of students who had made two or fewer school changes between 1998 and 2007 were from families with incomes below the poverty line, as compared with 26 percent of students who had experienced four or more school changes (Ashby, 2010). Mobility then tends to be highest in schools with high poverty concentrations, particularly in urban core areas.

For instance, one study found that about 25 percent of inner-city third graders were highly mobile, as compared with only about 15 percent of suburban and rural third graders (U.S. General Accounting Office, 1994). Within urban areas, transiency rates have been found to vary. Bruno and Isken (1996) reported average school-level transiency rates (the number of students entering and leaving a school over an academic year as a percentage of the school’s total enrollment) within the Los Angeles Unified School District at 42.6 percent. They further found rates to be variable and especially high at the elementary level, with some schools experiencing transiency rates as high as 95 percent. A 1992 study of the Chicago public schools noted an average student transiency rate of 36 percent (Bezruczko & Reynolds, 1992), while Kerbow (1996) reported that only 38 percent of sixth graders had attended the same school throughout their elementary years, and 13 percent had attended four or more schools during that period.

However, high student mobility is also common in economically disadvantaged rural areas as well. Schafft (2006), for example, examined five-year residential and educational histories from 22 families of students eligible for free or reduced-price lunch who had recently transferred into a rural district in upstate New York. On average, families had lived in six different residences over the five-year period with some respondents having lived in as many as 13 different places. In total, the children of these 22 households made 166 school changes over the five-year period, 92 percent of which were directly attributable to residential relocation. Most of the movement was between rural districts. This is consistent with work in other locations that has examined rural residential instability and school change (Fitchen, 1995; Foulkes & Newbold, 2005; Paik & Phillips, 2002).

In short, research on student transiency has tended to focus on its occurrence within urban locations or to use aggregated data gathered across the urban-rural continuum. However, despite the limited research on transiency in rural areas, mobility of low-income students should not simply be considered an urban...
phenomenon in either its incidence or its consequences for students and schools, particularly given consistently high nonmetropolitan poverty rates (Brown & Schafft, 2011). Moreover, as we discuss below, the smaller size of rural schools and districts means that the fiscal and human resource impacts of student transiency may be less easily absorbed than in a larger urban district.

Resource and Organizational Implications of Transiency for Schools and Districts

While there is now substantial empirical research on student transiency and its effects on social and academic outcomes, there is a pronounced lack of research on the organizational or fiscal impact of transience on schools as institutions. As previously discussed, a mobile student presents challenges to the classroom teacher, forcing teachers to reallocate their scarce resources to attend to the new student, or contend with behavioral difficulties that frequently accompany mobile children. The entrance or exit of a child from a school also initiates a sequence of administrative record-keeping steps (e.g., enrollment forms, eligibility analysis, course scheduling, and curricular alignment). These can result in increased staffing demands, reduced per-pupil resources, slowed school improvement, and a reduced capacity to engage in community-building efforts (Hartman, 2006). These examples are direct effects of mobility on educational resources. This section examines three additional topics related to educational financing, and their relationship to student transiency. These include school/district size, state aid calculations, and educational accountability.

Mobility, School/District Size, and Absorption Capacity

Transient students may represent liabilities to districts in terms of both increased fiscal costs and lowered test scores, particularly in economically disadvantaged areas. These are the districts that are least likely to have the resources or economies of scale to absorb large and unexpected costs affiliated with mobility (Paik & Phillips, 2002). The unexpected shifts in student population due to mobility represent a type of fiscal stress in these circumstances. As a superintendent from a rural New York district with an enrollment of about 650 explained (Schafft, 2005, p. 10):

There is increased pressure on school budgets. One sixth of our budget is targeted towards special needs kids. This year we had budgeted $100,000 for expenses associated with kids we anticipated would move into the district. We figured that we could expect four special needs kids to move into the district. We actually had ten move in, and we ended up needing to spend $250,000 to meet their needs, so we went $150,000 over our budget. We are a small district and so this was significant.

Fiscal stress occurs in this example due to the enrollment of high-cost students and associated instructional planning (e.g., staffing, facilities, and accommodations). Stress is also experienced through delayed reimbursement for those costs. Though state governments will reimburse districts for costs associated with the enrollment of special education students, these reimbursements are not immediate and rarely are they complete.

Although student transiency poses significant fiscal, administrative, and academic challenges across both urban and rural contexts, evidence suggests that larger, urban school districts may more easily absorb the challenges. In addition to district size and the enhanced ability of a larger urban district to absorb costs associated with individual students, some research in urban settings has suggested that much of the student movement is school-to-school from within the same district (Kerbow, 1996). In large urban districts there are more opportunities for mobile students to make school-to-school moves, without crossing district boundaries. Movements within the same school district and across schools pose administrative and record-keeping challenges, yet challenges are eased by circumstances such as aligned curricula, consistent academic calendars, instructional pacing, and centralized record-keeping. Such consistencies may not be enjoyed when mobile children cross district boundaries, which is frequently the case in small or rural school districts. Among rural districts, by contrast, there is often just one school for each set of grade levels. Therefore a child’s movement out of a school would likely mean entrance into a wholly different curriculum and course sequencing within a
Mobility, Enrollment Counts, and Impact on State Aid Formulas

Transience, as defined and described in this chapter, is not captured by conventional school finance formulae. Often aid formulas are simply not sensitive enough to adjust for the impacts of frequent student movement. Be it need-based aid or wealth-equalization formulas, a common feature in the calculus of state aid involves the development of an enrollment count for any one school district or school. Noting that enrollments vary over time, state aid policies commonly define enrollment as an average daily membership (ADM) and then enter this figure into a variety of allocation schemes (e.g., flat grants, foundation formulas, categorical, or need-based aid).

When enrollment is measured infrequently, state aid formulae and resource allocations are less likely to be responsive to the effects of mobile student populations. For example, in Vermont, ADM is calculated as a two-year average of enrollment counts taken on the 40th day of school. Some states, such as California, utilize two or more counts during the year to calculate their average. In general, student transiency, particularly in impoverished communities, reflects a “revolving door” where student exits and entrances are generally balanced. However, when enrollment counts are calculated infrequently, they become more biased to situations in which the churning of student populations are uneven.

State aid formulas may be particularly unwieldy in instances of small or rural districts with high mobility. For example, small districts often enter into contractual relationships with supervisory districts for vocational or special education services. Yet state aid formulas, to the extent they are based on average ADM figures, may not fully reimburse districts for sending and receiving transient students for some time. Moreover, if the students move prior to the calculation of ADM, it is possible that districts may not receive reimbursement at all for services they provide.

The state of Minnesota has addressed this problem by utilizing an individualized student enrollment system that minimizes the impact of mobility in their state aid formula. Minnesota’s foundation formula guarantees a fixed amount of revenue per pupil to ensure a minimally adequate education through a combination of state and local funds. Other categorical aid programs exist, but general state aid for the foundation plan constitutes the state’s largest funding stream. Though grade level pupil weighting is used in this system, at the heart of the formula is a weighted average daily attendance calculus (WADM) that is calculated using enrollment data from each day of the school year. State resources are allocated based on initial or expected WADM estimates. Yet, school districts also may submit enrollment updates at any time before the end of the school year and receive revenue adjustments. Minnesota’s finance formula is greatly aided by an individualized student information system that is more than 15 years old. The system allows state administrators and school district personnel to track students as they enter and exit schools to properly direct resources to schools for the costs incurred in providing services. Minnesota’s individual student database is akin to a spigot that allows resources to flow toward a school when students are in attendance at that facility.

What is absent from school funding formulae are enrollment weightings or indices that more directly measure student mobility. There are a variety of metrics that could be used such as a churning index, which could be the percentage of a student population that exited and entered the system after a defined date. Using a cumulative count of students involved in mobility would capture the costs associated with teaching outbound and inbound students, or administrative costs. A churning index may need to be measured at two points in time, given that most mobility-related moves occur shortly after the start of the school year, and again after January. A critical assumption in the creation of such a specific formulas is that it would be conceptually and empirically distinct from other adjustments, such as those for small schools or poverty.

Mobility and Performance Accountability

Mobility also confounds efforts to promote performance accountability within education. School-based performance accountability systems typically contain four main elements: clear educational goals; assessments of student achievement; methods to judge the effectiveness of schools in meeting the goals; and consequences,
particularly for low-performing schools (Stecher & Kirby, 2004; Figlio & Ladd, this volume). The issue of how to account for mobile schoolchildren is most relevant in test administration and in the overall analysis of school performance.

Regardless of the how a school’s performance is measured, schools face pressures to exclude mobile students from being tested and/or to exclude their test scores from school reports. For example, based on his examination of recent accountability models, Weckstein (2003) argues that current accountability policies can create incentives for administrators to exclude such students—many of whom are low-scoring—from schoolwide assessments. School officials might defend that action on the ground that the school often has limited time to teach such students. Further, given that mobile students may transfer out of the school before the testing at the end of the academic year, schools may have little incentive to devote much attention to them.

In some cases, accountability pressures are so strong that administrators attempt to manipulate measures of their school’s progress. For example, in case-study research in five New York State school districts, school administrators admitted to purposefully leading students at risk of dropping out into GED programs. Administrators described the practice as a mechanism to reduce dropout rates by labeling students as transfers (Sipple, Killeen, & Monk, 2004). In short, when schools are held accountable for changes in annual test scores at a particular grade level, administrators become quite sensitive to the sudden enrollment of students likely to upset those scores.

School mobility also complicates the task of measuring school progress, a central feature of current accountability policies. Schools are generally scored and comparatively ranked in one of two ways. In the first, schools are ranked based on either the level of the year-to-year change or in the aggregate performance of their students. Interestingly, state and local provisions differ in how they treat mobile schoolchildren for this purpose. Offenberg (2004) examined the relationship between school ranking and student mobility in Philadelphia and found that school rankings differ depending on the inclusion or exclusion of mobile schoolchildren. Figlio (2005) highlights this feature of school accountability policies for Florida school districts. He argues that the exclusion of test scores from high-mobility children may cause policymakers to overestimate the success of a school and lead to the misinterpretation of the success of a school in reaching the specified learning goals.

Accountability systems also rely on growth or value-added modeling to calculate school performance. In these multivariate regression models, changes in student performance at the individual level are tracked over time. The goal of the more sophisticated of these approaches is to parse out known factors that contribute to student performance from those attributable to the school itself. In practice, though, as Figlio (2005) notes, it is not possible to control for the complete set of background characteristics associated with student performance (e.g., wealth, race, and mobility), and thereby to disentangle factors contributing to school performance. Unmeasured student, family, and community characteristics—including student transiency—represent a major shortcoming of the growth models that are attempting to isolate the contribution schools make to student learning.

Beyond the calculation of student and school performance, student transiency invokes more fundamental questions about educational accountability itself. How much time is necessary for an educator or a school to have an impact on children’s educational outcomes? Which part of the educational system (e.g., teachers, schools, or districts) ought to be held accountable if a child’s enrollment has been too brief or too close to the date of comprehensive testing in any one school? These are important questions that are not resolved simply by removing mobile students from testing opportunities, or by implementing policies that remove their scores from aggregate measures.

**Conclusion: Some Direction for Policy Interventions**

While schools have relatively little control over the household-level economic circumstances that place children at higher risk of frequent school change, school-based interventions may offer possibilities for ameliorating the impact of such moves.

At the conclusion of one of the most robust empirical studies of the deleterious effects of frequent mobility on student achievement, Hanushek et al. (2004, p. 1746) state:
A policy challenge ... is to devise schooling approaches that mitigate the academic losses due to school turnover. Whether, for example, such things as more standardized curricula, specialized transition, and remedial programs for entering students, or more careful classroom placement of new students, could help, remain open questions.

While a variety of programmatic and policy options may be considered, authors writing about student mobility cite large demographic variability among homeless and mobile students as a reason for flexible interventions (see Obradović et al., 2009).

Though empirical evaluations of interventions for highly mobile students are limited in number, qualitative research supports school-based interventions focused on heightening school attachment among students. School attachment is generally defined as a sense of connectedness and commitment that results from a perceived sense of caring from peers and teachers (Libby, 2004; Wilson, 2004). As school attachment increases among students, positive achievement and pro-social outcomes become more likely, including improved academic achievement, reduced delinquency rates, and decreased likelihoods of harmful health practices (Rogers, 2004).

Parental involvement in building student attachment in the context of mobility has been a focus of scholarship since the late 1980s (see Jason et al., 1989). In a recent and promising study, Fiel, Haskins, and Lopez-Turley (2013) report on a cluster-randomized field experiment among elementary schools in San Antonio, Texas, and Phoenix, Arizona. This intervention, which sought to improve relationships among students, families, and schools, showed reduced mobility rates among black students. Literature on parental participation in school programs stresses the development of strong collaborative value systems among parent-professional partnerships (Lake & Billingsley, 2000; Muscott, 2002). Strong partnerships can ameliorate barriers to optimal parent participation in schools (Furney & Salembier, 2000), improve attention to diverse race and ethnic subgroups in schools (Callcott, 2003; Kalyanpur & Harry, 1997), as well as improve individual and systemic educational service provision for children with exceptional needs (Lopez, Scribner, & Mahitivanichcha, 2001; Rao, 2000).

Strengthening parental engagement clearly should be considered part of the portfolio of intervention strategies, but many challenges remain. First, parents of low-achieving, highly mobile students are likely themselves to have had mixed academic experiences. Their own negative academic experiences may make them hesitant partners in such efforts. Secondly, the social and economic circumstances of mobile, low-income parents, including uncertain employment and childcare arrangements and potentially limited access to transportation and communication, pose additional barriers. Finally, the nature of high mobility in and of itself means that parents may not be residentially stable long enough to develop effective partnering relationships with schools.

Other interventions supersede the institutional level of the school, but nonetheless have significant implications for addressing transiency. For example, scholars have evaluated federal and state efforts to move low-income households located in socially, economically, and racially segregated neighborhoods into more heterogeneous and middle-class communities. Initial evidence suggests that policies such as housing vouchers can work to stimulate residential homeownership and desegregation among low-income and minority households (Clark, 2005; Johnson, Ladd, & Ludwig, 2002; Rosenbaum et al., 2002). These policies are hopeful for their apparent embracing of a richer and more complex understanding of population movement, housing, and poverty dynamics.

Regardless of the innovative interventions that may be devised, causes of childhood mobility are multidimensional and cannot be meaningfully disentangled from the varied dimensions of poverty and social inequality that are woven into the fabric of contemporary American society (Evans, Yoo, & Sipple, 2010). Given current educational mandates and institutional realities, it is appropriate that education policy facilitate the social and academic adjustment—and ultimately the success—of high-need, at-risk youth who are quite literally “moving targets.” Therefore, educational researchers make important contributions by generating knowledge about the causes and consequences of student mobility and the potentials of various intervention models. We argue, however, that fully grasping the problem of student transiency requires widening the analytic focus beyond the school walls and beyond the indicators of learning achievement. Absent a wider analytic lens, educational researchers may miss the true root of the problem: many American adults and children lack the basic means to lead socially and economically productive lives.
Notes

1. See Title X, Part C, Section 725(2) of the No Child Left Behind Act of 2001.

2. According to current (4/2014) enrollment counts, 80,254 children are enrolled in Pre-K through Grade 12 Department of Defense schools, located within the United States and overseas (see http://www.dodea.edu/datacenter/index.cfm). Another 600,000 children of military personnel are enrolled in U.S. K–12 public schools located near military bases (Military Family Resource Center, 2001).

3. However, the income level does not factor in the multiple benefits that accrue from military employment, such as housing allowances, that offset the lower pay scale for enlisted personnel.


References


Evans, G. W., Yoo, M. J., & Sipple, J. (2010). The ecological context of student achievement: School building quality effects are exacerbated by high levels of student mobility. *Journal of Environmental Psychology, 30* (2), 239–244.


Bridging the High School-College Divide

THOMAS R. BAILEY AND MELINDA MECHUR KARP

High schools and colleges are financed, managed, and regulated by separate systems and agencies. This arrangement has resulted in inefficiencies and inequities as students move between the two sets of institutions —reducing retention and completion in college and creating inequities in enrollment and college completion patterns. As college enrollments have grown (Kane, 2004; Snyder & Dillow, 2012) and the economic value of college degrees has risen over the last two decades (Baum, Ma, & Payea, 2013), policy-makers and educators have sought to improve the alignment between high school and college and thus to facilitate the process by which students move between the two sets of institutions. This chapter reviews three strategies designed to promote these goals.

The first is the movement to align high school and college curriculum and standards. This is an explicit attempt to get states to minimize inefficiency by creating a coherent pathway that spans the secondary and postsecondary sectors. It requires state-level regulatory changes and collaboration among a wide variety of institutions.

The second approach, dual enrollment, enables high school students to earn college credit while still in high school. It is designed to facilitate the transition to college by giving students a college experience prior to enrolling in postsecondary education, thereby preparing them to navigate the high school-college divide more successfully. This strategy requires collaboration between high schools and colleges. It is generally carried out through specific relationships between individual institutions, sometimes within a framework established by the state.

Finally, remediation provides instruction for students entering college with academic skills so weak that the students are not prepared for college-level work. Remediation, also known as "developmental education," is, partly, an attempt to compensate for mis-alignment and the failure of the K–12 system to prepare students for college. Colleges can carry out remediation on their own, without collaboration from high schools, but increasingly, colleges and high schools are working together to reduce the need for remediation or take steps in high school when students appear to be on track for remediation if they go to college.

The long-standing organizational separation of the high school and college systems has resulted in misalignment between the completion of high school and the beginning of college. For example, although a majority of students arriving at community college are referred to remediation (Attewell, Lavin, Domina, & Levey, 2006; Bailey, Jeong, & Cho, 2010), almost all of those students are high school graduates. Clearly a high school diploma is not taken by college faculty and counselors to indicate full preparation for college.

One result of this misalignment is that students are misinformed about what they must do to be ready for college or what to expect when they get there (Achieve, 2006; Venezia, Bracco, & Nodine, 2010; Venezia, Kirst, & Antonio, 2003). Students who arrive misinformed are more likely to have problems in college, to have less successful experiences, and to leave college before reaching their goals or completing degrees. This is one reason why the growth of college enrollments has outpaced college completion (Turner, 2004).

The separation between high school and college also reinforces inequities with respect to socioeconomic status (SES) in access to college and college completion. High SES students have many advantages over low SES peers in navigating the transition from high school to college. They tend to be in schools that are more effective at preparing them substantively for college; they have access to more extensive and effective counseling services; their parents, relatives, and friends are much more likely to have knowledge about the process; and they have more resources with which to purchase material and services to help them find and enroll in an appropriate college. A more organized and aligned transition process could help lower-income students overcome some of these disadvantages.
High School-College Alignment

One strategy to help students move from one educational sector to another is to minimize the separation between those sectors. Better-aligning the curriculum, standards, assessments, and expectations in high schools and colleges could improve student transitions by creating a coherent pathway. Thus students who graduate from high school will know that they are prepared for college.

Over the past decade, educators have introduced a variety of strategies to create a coherent path from high school to college. These efforts seek to ensure that the curricular and academic performance standards are the same for exiting high school as they are for entering college. They also seek to ensure that all students meet college-ready standards prior to high school graduation, thereby ensuring that they are able to smoothly enter college should they wish to do so.

The most prominent effort to align high school and college is the Common Core State Standards (CCSS). Finalized in 2010, the CCSS are an attempt to create a consistent framework outlining what students should know to be ready for college and careers. Though voluntary, 45 states had adopted the standards as of July 2012. Two sets of assessments are currently being developed to assess whether or not students have met the standards. States are beginning to use the CCSS to guide curriculum and instruction, and use of CCSS-aligned assessments will begin in 2014.

As the Common Core is just being deployed, it is too early to determine whether it has improved high school-college alignment. Some authors have expressed concern that, despite being aimed at college readiness, the postsecondary sector was not systematically involved in the development of the standards (Barnett & Fay, 2013). However, there is reason to believe that the CCSS will decrease the divide between the two education sectors; at least six national initiatives explicitly focus on bringing higher education into the Common Core conversation in order to further promote the goal of intersector alignment. Many states are also working to more explicitly link the CCSS to college readiness, for example by seeking to use a passing score on a Common Core assessment as a qualification for placing out of developmental education (McMurrer & Frizzell, 2013).

Early college readiness assessments are a second approach to improving cross-sector alignment. States and colleges use a variety of assessment tests to determine whether incoming college students are ready for college credit courses. However, these assessments are typically given to students when they enter college—too late to prevent their entry into developmental education. By offering these assessments to high school students (usually in 11th grade), early college readiness assessment initiatives seek to inform students about their readiness for college-level work prior to their actual matriculation into postsecondary education, when they still have time to address any deficiencies.

Thirty-eight states offer early college readiness assessments, though in some cases this is via local rather than statewide initiative (Barnett, Fay, Bork, & Weiss, 2013). These strategies are relatively new, and little research has been conducted on their impact. Those studies that have been conducted find small but positive outcomes for participants (Howell, Kurlaender, & Grodsky, 2010; Jobs for the Future, 2012; Kerrigan & Slater, 2010).

Telling students whether or not they are ready for college via early college readiness assessments may help inform students that they need extra preparation but does not ensure that they become college-ready. Therefore, some states and districts have developed a third alignment strategy, transition courses, that seek to intervene with high school students assessed as unprepared for college. Transition courses (or transition curricula) are offered to students during their junior or senior year to address their lack of preparation for college-level work. These courses, often developed jointly by high schools and colleges, provide a closely aligned curriculum that directly addresses the demands of college coursework thus potentially reducing the need for remediation.

Though transition courses have existed on a small scale for many years, their use as a comprehensive transition strategy at the district and state level is quite new (Barnett, Fay, Trimble, & Pheatt, 2013). Most initiatives remain locally developed and offered; though these courses exist in 29 states, only eight have statewide transition course initiatives (Barnett, Fay, Bork, & Weiss, 2013). While promising, these activities are too new to assess, with no current studies evaluating student outcomes from participation (Barnett, Fay, Trimble, & Pheatt, 2013).
Dual Enrollment

A second strategy, dual enrollment (also called dual credit or concurrent enrollment), does not attempt to ameliorate the divide between high school and college, but seeks to help students more successfully navigate that divide. It does this by enabling high school students to take college courses and, sometimes, earn high school credit simultaneously (Bailey & Karp, 2003; Western Interstate Commission for Higher Education, 2006). In doing so, students earn college credit and are exposed to college-level coursework, as well as the expectations and norms of college courses. As a result, they may be more prepared to navigate postsecondary education successfully after high school graduation.

Dual enrollment has grown rapidly over the past decade. Recent research by the National Center for Education Statistics found that nearly two million public high school students took a dual credit course during the 2010–11 school year; this represents an annual growth rate of nearly 7 percent since 2002–03 (Shapiro, Dundar, Ziskin, Yuan, & Harrell, 2013; Thomas, Marken, Gray, & Lewis, 2013). Eighty-two percent of public high schools had students enrolled in dual credit courses in 2010–11 (Thomas et al., 2013). Forty-seven states and the District of Columbia now have laws regulating dual enrollment (Education Commission of the States, 2013).

In many schools, dual enrollment attracts young people who are not among the most academically successful students, including many who in the past would not have been considered college bound. Thus many dual enrollment students have B averages and have not taken the most rigorous “college prep” sequences (Hughes, Karp, Fermin, & Bailey, 2006). Additionally, many states fund dual enrollment such that it is cost-free to students and their families, allowing them to save money by earning free college credit (Borden, Taylor, Park, & Seiler, 2013; Education Commission of the States, 2013). Therefore, dual enrollment is a popular strategy to increase postsecondary success for lower-achieving and low-income students, who are often less successful than their more-advantaged peers.

Dual enrollment can help students navigate the fragmented high school-to-college transition for several reasons. First, students may benefit from access to rigorous, college-level courses and free college credit. By accruing college credits early, they may develop academic momentum—building a “nest egg” that can help propel them to further college success. Academic momentum has been associated with college persistence (Adelman, 2006; Swanson, 2008). Second, many educators believe that teacher and parent expectations are crucial determinants of student performance and that students will respond if their teachers expect them to perform at a higher level. In other words, treating young people like college students will raise their academic performance.

Third, and perhaps most powerfully for low-income and first-generation college students, by exposing students to a college-level experience, dual enrollment both shows them what to expect in college and gives them the opportunity to practice college behavior. The hypothesis is that students with these early experiences will not be surprised—and possibly discouraged—when they do enroll in college. Better preparation will, according to this view, reduce early attrition from college and promote longer-term college success (Karp, 2012; Wolk, 2005).

There is a growing body of research indicating that dual enrollment positively influences participants’ college outcomes. Qualitative research finds that dual enrollment participants learn study skills and other habits related to college success, as well as gain the opportunity to practice these skills and strive to meet other collegiate expectations (Foster & Nakkula, 2005; Karp, 2012; Nakkula, 2011). Quantitative studies that control for measured student characteristics find that dual enrollment participants graduate from high school (Cowan & Goldhaber, 2013; Karp, Calcaigno, Hughes, Jeong, & Bailey, 2007; Rodriguez, Hughes, & Belfield, 2012) and enroll in college at greater rates than their non-participating peers (Cowan & Goldhaber, 2013; Karp et al., 2007; Rodriguez et al., 2012; Speroni, 2011a; Struhl & Vargas, 2012).

Dual enrollment participation is also related to successful outcomes once students enter college. Studies have found that, compared to nonparticipants, dual enrollment students have higher college grade point averages (Allen & Dadgar, 2012; Eimers & Mullen, 2003) and greater rates of persistence into their second year of college (Eimers & Mullen, 2003; Struhl & Vargas, 2012; Swanson, 2008), rates of credit accrual (Cowan & Goldhaber, 2013; Karp et al., 2007; Michalowski, 2007; Rodriguez et al., 2012) and likelihood of degree completion (An, 2013; Struhl & Vargas, 2012). Importantly, these outcomes hold true for all kinds of students, including those in...
career and technical education programs (D’Amico, Morgan, Robertson, & Rivers, 2013; Karp et al., 2007; Rodríguez et al., 2012; Struhl & Vargas, 2012) and those from low-income families and who are the first in their families to attend college (An, 2013; D’Amico et al., 2013; Rodríguez et al., 2012; Struhl & Vargas, 2012).

Dual enrollment programs can be conducted under a variety of arrangements, for example, either at the high school or on a college campus. Not all formats have equal outcomes. For example, dual enrollment programs held on college campuses may have a greater impact on student outcomes than those in which courses are held on high school campuses (D’Amico et al., 2013; Speroni, 2011a). Other studies find that the type of dual enrollment course and the rigor with which it is taught influence whether or not participation impacts student outcomes later on (D’Amico et al., 2013; Allen, 2010; Speroni, 2011b). Furthermore, although educators have worked to attract a wide range of students, so far white, female, and higher-SES students disproportionately participate in dual enrollment (Pretlow & Wathington, 2013).

Moreover, it is important to note that dual enrollment students are likely to be more motivated than the broader population of high school students. Most dual enrollment programs have entry criteria (Borden et al., 2013), and participants are choosing to take a rigorous course. Most of the studies that we have cited address this problem primarily by controlling for measured characteristics so were unable to account for unmeasured factors such as motivation, although Speroni (2011b) used regression discontinuity, a stronger causal methodology.

There are some very recent random assignment studies with stronger causal conclusions on a very ambitious dual enrollment–based strategy called Early College High Schools (ECHS). This is a more comprehensive strategy based on a more thorough redesign of the high school enabling all students to earn one to two years of transferrable college credit while completing the requirements for a high school diploma. Early college high schools also provide students with comprehensive supports to help them develop the academic, social, and behavioral skills necessary for college success.

A rigorous study of 10 early college models found modest positive effects for high school completion but stronger positive effects for college enrollment (Berger et al., 2013). Twenty percent of the students in the study actually completed a college degree by the end of high school. The schools in this study chose students through lotteries and therefore the researchers were able to compare outcomes for lottery winners and losers. A random assignment study of early colleges in North Carolina also found positive effects on high school completion and college enrollment (Edmunds, Unlu, Glennie, Smith, & Bernstein, 2013). The even more ambitious Pathways in Technology Early College High School (P-TECH) is a six-year program that fully integrates high school and the first two years of college—students earn both an associate’s degree and a high school diploma. It attracted a great deal of attention when President Obama visited the school in 2013, but it is too new to allow any evaluation (Foroohar, 2013).

Therefore a growing body of evidence suggests that dual enrollment experiences can help students navigate the transition from high school to college. This may not be surprising since the strategy at least partially integrates the two sets of institutions by having students engage in college-level activities while still in high school. Very recent rigorous random assignment research also suggest that reforms of high school with more thorough incorporation of college experiences—the ECHS—does have positive effects on college enrollment.

Developmental Education and Services for Academically Underprepared Students

Once students arrive in college, many find that they are not prepared for college-level coursework. A third strategy, called remediation or developmental education, aims to help these students become college-ready once they have already entered postsecondary education. At all community colleges and most less-selective four-year institutions, students are assessed when they arrive for registration and referred to remediation if their assessment scores fall below an established cutoff. About one-half of all entering college students and 68 percent of entering community college students take at least one remedial course within six years of first enrollment (National Center for Education Statistics, 2013). Many of these students enroll in more than one remedial course, either in one subject or in multiple subjects (Attewell et al., 2006; Bailey et al., 2010).
Approximately one-Fifth of entering students are referred to sequences that require at least three courses in order to prepare for college-level math (Bailey et al., 2010).

Over the last five years, there has been a tremendous amount of reform of the assessment and placement process, and developmental education itself. Several states including Connecticut, Florida, North Carolina, Virginia, Tennessee, and Colorado have initiated significant and sometimes controversial redesigns of their developmental education systems. The nonprofit organization Complete College America is working with a network of 33 states and the District of Columbia to promote major changes in remediation (http://www.completecollege.org/alliance.html), and the topic was featured at the January 2014 White House Summit on College Admissions and Success (White House, Office of the Press Secretary, 2014).

Much of this extensive activity is motivated by research findings that suggest that remediation as it is has been traditionally organized and taught has not been successful in bridging the gap between high school and college and may actually create barriers. The charge of developmental education is clear: to build up the skills of academically underprepared students so that they may be successful in college-level work and progress quickly to a credential that will advance their position in the marketplace. Unfortunately, as the system currently stands, this goal is rarely met. Fewer than half of students successfully make it through the sequence of courses to which they are referred (Bailey et al., 2010). Many students fail to complete their sequence of remedial courses without actually failing or withdrawing from a course—they drop out between the courses in the long sequences to which they are referred (Bailey et al., 2010). Low-income, minority, and first-generation college students are all overrepresented in these negative outcomes associated with developmental education (Perry, Bahr, Rosin, & Woodward, 2010).

Evaluations of the effectiveness of remediation that have relied primarily on regression discontinuity designs have found mixed outcomes at best. Most of these studies suggest that the traditional system of remediation does not improve student skills enough to increase their chances, on average, of passing college-level courses or completing degrees or certificates (Calcagno & Long, 2008; Clotfelter et al., 2013; Dadgar, 2012; Martorell & McFarlin, 2011; Scott-Clayton & Rodríguez, 2012; Xu, 2013). Some studies have found some positive effects for some groups of students. Bettinger and Long (2009) conclude that remediation has positive effects overall. This was a sample of students who went to four-year colleges. In an article focused on full-time community college students, they found positive effects for math remediation but no effects for reading (2005). The sample in this study was limited to the relatively small share of community college students who took the ACT and who declared their intention to transfer. Boatman and Long (2013) also found some positive effects but just as many negative effects for remedial courses for students in longer sequences.

Thus traditional developmental education did not serve as an effective bridge between high school and college, or at least the college-level program. Most studies find that it does not improve outcomes yet it increases the costs to the student and the funders of higher education, and delays their progress towards graduation. Indeed, Scott-Clayton and Rodriguez (2012) concluded that remediation diverted students away from college-level programs, a conclusion confirmed by Clotfelter and his associates (2013).

Problems with developmental education start with the assessment system designed to find out who really needs remediation. Recent research has shown that the standardized assessment instruments commonly used to place students in remediation do a poor job (Fuenmayor, Hetts, & Rothstein, 2012; Scott-Clayton, Crosta, & Belfield, 2014). In one study, more than a fifth of students placed into math remediation could have earned a B or better in a college-level math course; as many as 40 percent could have earned a C (Scott-Clayton et al., 2014). On the other hand, some students who are placed into college-level courses do not have adequate skills and fail. Another problem with the assessment and placement system is that students are often poorly informed about the assessment tests and their consequences (Venezia et al., 2010).

Several researchers have proposed placement system models that go beyond primary reliance on standardized assessments, using a variety of student skill measures such as high school GPA and noncognitive assessments (Conley, 2005; Sawyer, 1996; Sedlacek, 2004; Levine-Brown, Bonham, Saxon, & Boylan, 2008). Simulations suggest that assessments using GPA instead of or in addition to the traditional assessments would more accurately predict who would be successful in a college-level course (Scott-Clayton, Crosta, & Belfield, 2014).

In the face of research showing limited effectiveness of developmental education, educators have designed a variety of reforms (Rutschow & Schneider, 2011). But most of these reforms have addressed one element of the
developmental education process, have provided only a short-term intervention (often lasting just one semester), and have involved limited numbers of students (Barnett et al., 2012; Boatman, 2012; Visher, Weiss, Weissman, Rudd, & Wathington, 2012). Rigorous research is beginning to emerge that measures the effectiveness of these types of programs.

For example, learning communities have been a popular reform for 10 to 15 years (Bailey & Alfonso, 2005). In learning communities, students are usually placed in a cohort where they take a paired college-level and remedial course. A random assignment study of learning communities in six colleges found modest, positive effects that faded over time (Visher et al., 2012). Similarly a rigorous study of summer bridge programs designed to provide some assistance to students with low assessment scores during the summer between their senior year in high school and their first year in college found only modest and temporary effects (Barnett et al., 2012).

Given the poor outcomes for students in long sequences of remedial course, a variety of models are designed to shorten the time required to get through developmental education. Modularized strategies are built on new assessments that identify specific student weaknesses. Sometimes it does not take a full semester to address those problems, so students are provided instruction only in their areas of weakness (Bickerstaff, Fay, & Trimble, in preparation). Other strategies used at the Community College of Denver and Chabot College compress two or more developmental courses into one semester. These have been able to place more students into initial college-level courses, also without lowering passing rates (Edgecombe, Jaggars, Baker, & Bailey, 2013; Edgecombe, Jaggars, Xu, & Barragan, in preparation). An important element of these strategies is that developmental students practice the same assignments they would receive in college-level courses, and are provided with targeted feedback and additional instruction to help them perform more strongly on these assignments, thus giving students a better preparation for those courses.

Since poor outcomes for developmental education also appear to stem from misalignment between the content of remedial and college-level courses (Jaggars & Hodara, 2013), a variety of reforms have been designed to align the two more closely. One encouraging program—Washington State’s I-BEST program—contextualizes remedial instruction with the substantive content in occupational programs. The remedial courses must be part of a “career pathway” which leads directly to a credential and to jobs that are in demand in the local labor market. Basic skills occupational students in colleges that adopted I-BEST were about 10 percentage points more likely to earn college-level credits and 7 percentage points more likely to earn a certificate within three years (Zeidenberg, Cho, & Jenkins, 2010). This type of approach is particularly popular in shorter-term programs leading to occupational certificates.

Several programs now are designed to align the math content of remediation with specific student goals. These strategies all start from the observation that traditional developmental math has been designed to lead to college-level algebra as a preparation for calculus, but most students do not need calculus and few ever take it. So these programs, referred to as Statway, Quantway, and Mathway, target students who are headed for programs that do not require calculus and prepare them for college-level statistics or quantitative reasoning courses. They also use college-level pedagogy and design the courses to be much more closely integrated with the college-level content. Initial descriptive outcomes for Statway show very large increases in enrollment and completion of college-level courses (Van Campen, Sowers, & Strother, 2013).

The so-called co-requisite model of remediation is also becoming increasingly popular. In this approach, students are enrolled simultaneously in both a college-level course and a specially designed remedial course aligned with the college-level course. The best-known example of this is the Accelerated Learning Program (ALP) at the Community College of Baltimore County. ALP students were 33 percentage points more likely to enroll in and pass the first college-level English course than were non-ALP students (Cho, Kopko, Jenkins, & Jaggars, 2012). ALP and similar programs have now spread to dozens of colleges.

The most ambitious program is the City University of New York’s Accelerated Study in Associate Programs (ASAP) initiative, which embeds developmental education in a broader structure of curricular pathways and support services. ASAP students are encouraged to take their developmental courses early in their college careers; they also enroll in block-scheduled classes with other ASAP students and a targeted student success course. Early results of the ASAP program are promising. A random assignment study found that after two years, ASAP students in need of remediation were more likely to enroll in college and remain enrolled than similar, non-ASAP students; accrued college credits at a greater rate; and were more likely to graduate
Though undoubtedly not all of these benefits accrue from changes to developmental coursework, it is important to note that by moving students through developmental education more quickly, the ASAP approach enables students to progress into college credit-bearing coursework and, ultimately, graduate.

The field of developmental education is changing rapidly. Research in just the last five to seven years has shown that the traditional system of assessment and instruction is not effective and may create additional barriers to students who already face many challenges. This has generated a wide variety of encouraging initiatives and reforms in many states. This suggests that in five years, the services for students with weak academic skills will look very different than they do, even today.

A theme that runs through most of these reforms is that they foster a stronger relationship between developmental education and the content and pedagogy of the college-level programs. I-BEST contextualizes the basic skills instruction with the content of occupational programs. Compressed programs such as those at the Community College of Denver and Chabot College move students as quickly as possible into college courses and emphasize the pedagogies used in those courses. Statway and the other math reforms provide different pathways through remediation based on students’ desired college programs. Co-requisite models such as ALP actually enroll the remedial students into the college course. ASAP is a comprehensive program with a strong focus on guiding a student through their entire program, and the remedial services are closely coordinated into that process. Research on all of these reforms has shown positive outcomes, although the studies vary in the strength of their causal conclusions. If, as research indicates, traditional remediation did not serve as an effective bridge between high school and college, the emphasis on connecting services for students who have weak academic skills to their college programs of study would in principle significantly strengthen the high school-to-college transition.

But much remains to be done. Positive results for some of these programs have been established through correlational or descriptive analyses. Moreover, research has also shown that positive effects of short-term programs that last only a semester or two tend to fade over time. Even encouraging reforms of remediation are not powerful enough to influence overall graduation rates (Rutschow et al., 2011; Quint, Jaggars, Byndloss, & Magazinnik, 2013). The next step is to figure out how to combine these individual reforms into comprehensive programs, like CUNY’s ASAP program, that carry the student through to graduation. Then developmental education will not only help bridge students from high school to college, but from high school to college graduation.

**Conclusion**

Research on these three strategies has grown rapidly in the last five years and has had a widespread influence on policy and practice. This is particularly true of developmental education, which is going through a period of fundamental reform. This is somewhat less true of dual enrollment, although our understanding is growing, especially for the more comprehensive models—Early College High Schools. We have some knowledge of the effects of early assessment and transition courses. Widespread efforts to promote alignment between high school completion and college entry are now focused around the development of the Common Core, and it is too early to judge the effectiveness. We do know that the coordination of the involvement of the two sectors in the development of the Common Core and related assessments has not been easy.

A straightforward research agenda flows from the research presented in this chapter. First, we need more rigorous evaluations of the new developmental education reforms, especially co-requisite models that are spreading rapidly. What policies would work best for developmental students with very weak skills—the type of students who are now referred to three levels of remedial courses—remains a poorly understood question. The multiple sequences of courses that are typical today is clearly not working, but popular reforms such as the co-requisite strategies seem unlikely to work for the weakest students.

Rigorous research on the effect of dual enrollment courses is still scarce. One important question is whether it makes a difference whether the courses take place at the high school or on the college campus. The former is much easier to scale to larger numbers of high school students, but the more ambitious early college models may have more chance of having a more profound effect. In general, we need further experimentation and
evaluation about how to integrate or blend late high school and early college. The P-TECH model completely integrates them by creating a six-year high school that ends with an associate’s degree. This creates a strong bridge between high school and college, but outcomes have not been studied.

It is of course too early to study the effects of alignment based on the Common Core. Nevertheless, researchers need to track the development and utilization of the assessments and begin to conceptualize how to think about and measure the implications of this broad reform.

We have not discussed the cost of these reforms. Higher education remains a good investment so it makes sense to increase spending on efforts to get most students successfully into and through college. Nevertheless, colleges are not likely to enjoy significant increases in resources so cost must be an issue. Indeed, some of the successful or encouraging programs that we have discussed such as ALP, ASAP, and I-BEST all cost more per student, although evaluators point out that they often cost the same or less for successful outcomes (Jenkins, Speroni, Belfield, Jaggars, & Edgecombe, 2010; Levin & Garcia, 2013; Zeidenberg et al., 2010).

In the end, it may be most productive to think about integrating these three strategies rather than to continue to develop and study them as separate approaches to bridging the high school/college divide. Research has suggested that focused reforms that address one element of the process of moving from high school into college-level programs and through to graduation often have modest or even no measurable impacts, but that does not necessarily mean that combining those strategies into a more comprehensive reform will not have a strong positive effect. Better alignment and college-like experiences in high school may reduce the need for remediation. Even for students who continue to have weak skills as they enter college, they will have a better foundation to allow them to succeed in remedial programs designed to speed and solidify their attachment to college-level programs of study.

In the past, research on the link between high school and college has been hampered by lack of data, but now 44 states link high school and postsecondary data (Data Quality Campaign, 2013). Comprehensive reforms that involve whole institutions or even combinations of institutions are much more difficult to evaluate than well-defined and discrete programs. But if we want more profound change, it is unlikely to be achieved through small programs. It is up to researchers to figure out how to measure such strategies, rather than focusing on easier-to-study reforms which are likely to have at best modest effects.

Society has reached a consensus that every student needs some postsecondary education, yet we continue to live with an institutional structure that emerged when only a small percentage of students went on to college. The basic structure of those institutions is unlikely to change, but there are many initiatives that are working to draw together and align high schools and colleges. Research has already had a profound effect on that process and can continue to support and strengthen the trend.

Notes

1. Better alignment and indeed more effective high schools will not eliminate the need for some type of developmental education in college, because many college students are adults returning to college after several years in the labor market or they are immigrants who attended high school abroad and who at least need help with English.

2. Among students who graduated from high school in 2011, rates of enrollment in college the following fall were 30 percent lower for low-income students (52 percent) than for high-income students (82 percent) (Aud et al., 2013). For those who enter college, completion is also highly related to SES. For example, eight years after high school graduation, just over 20 percent of students from the lowest SES quartile have completed a degree, while almost 70 percent of those from the top quartile of the SES distribution have a degree (Lauff & Ingels, 2014).


4. Dual enrollment differs from the College Board’s Advance Placement (AP) program in that AP students do not earn credit by taking the AP course, but rather by scoring above a cutoff score on the AP exam and by having colleges accept those scores for credit. Dual enrollment students earn credit by completing the course, as do college students in college.

5. To ease exposition we use the terms "developmental education" and "remediation" interchangeably. Although practitioners and professionals sometimes interpret the terms somewhat differently, both refer to services provided to students who are judged to be ill-prepared for college-level coursework.
References


Borden, V.M.H., Taylor, J. L., Park, E., & Seiler, D. J. (2013). *Dual credit in U.S. higher education: A study of...


Karp, M. M. (2012). "I don’t know, I’ve never been to college!” Dual enrollment as a college readiness strategy. 


Economic Research.


Contributors

General Editors

Helen F. Ladd is the Susan B. King Professor of Public Policy Studies and Professor of Economics at Duke University’s Sanford School of Public Policy. She edited Holding Schools Accountable: Performance Based Reform in Education (Brookings, 2006) and was co-chair of a National Academy of Sciences Committee on Education Finance from 1996 to 1999. She is co-editor of the first edition of this Handbook, the author of many articles on U.S. education policy, and the co-author of books on school choice in New Zealand and school reform in post-apartheid South Africa. She is a past president of the Association for Public Policy Analysis and Management (APPAM) and a member of the National Academy of Education.

Margaret E. Goertz is Professor Emerita of Education Policy and Senior Researcher, Consortium for Policy Research in Education, in the Graduate School of Education at the University of Pennsylvania. Her research focuses on state and federal education finance and governance policy. She is co-author of books on school finance reform in New York and New Jersey and co-editor of School-Based Financing (1999). She has also written extensively about state education reform policies, state and federal programs for special needs students, and the implementation of Title I. She is a past president of the Association for Education Finance and Policy (AEFP) and a recipient of AEFP’s Outstanding Service Award.
Section Editors

**Henry M. Levin** is the William Heard Kilpatrick Professor of Economics and Education and Director of the National Center for the Study of Privatization (NCSPE) at Teachers College, Columbia University and the David Jacks Professor Emeritus of Education and Economics, Stanford University, where he served on the faculty from 1968 to 1999. A specialist in the economics of education, he is a recipient of AEFP’s Outstanding Service Award.

**Susanna Loeb** is the Barnett Family Professor of Education at Stanford University, director of the Center for Education Policy Analysis, and Co-Director of Policy Analysis for California Education. Her research addresses policies affecting educator workforce development, school finance, and school governance. She is a past president of AEFP.

**David H. Monk** is Professor of Educational Leadership and Dean of the College of Education at The Pennsylvania State University. A past president of AEFP, he was the inaugural co-editor of the Association’s journal, *Education Finance and Policy*. He is also a recipient of AEFP’s Outstanding Service Award.

**Andrew Reschovsky** is a Fellow at the Lincoln Institute of Land Policy in Cambridge, MA and a Professor Emeritus of Public Affairs and Applied Economics, University of Wisconsin-Madison. He has conducted research and published articles on tax policy, educational finance, and intergovernmental relations, and serves as an advisor to the government of South Africa on local government fiscal reform and educational finance. He received APPAM’s Steven D. Gold award in 2011 in recognition of his contributions to state and local fiscal policy.

**Jennifer King Rice** is Professor of Education Policy and Associate Dean for Graduate Studies and Faculty Affairs, College of Education, University of Maryland. Her research draws on the discipline of economics to explore education policy questions concerning the efficiency, equity, and adequacy of U.S. public education. Her current work focuses on teachers as a critical resource in the education process. She is a past president of AEFP.

**Leanna Stiefel** is Professor of Economics and Education Policy at the Wagner Graduate School and the Steinhardt School, New York University. Her research on education finance and policy has been published as books, including *The Measurement of Equity in School Finance* (with Robert Berne) and *Measuring School Performance and Equity* (co-edited), and in numerous journals, including the *Journal of Education Finance and Policy*. A past president of AEFP, she received its Outstanding Service Award in 2005.

**Ron Zimmer** is an Associate Professor of Public Policy and Education in the Peabody College, Vanderbilt University. He has conducted research on school choice, school finance, and accountability policies funded by an array of foundations and the Institute of Education Sciences. He is currently a co-editor of *Educational Evaluation and Policy Analysis*, an editorial board member of *Economics of Education Review*, and a member of the Board of Directors of AEFP and the School Choice conference.

Authors

**Melinda Adnot** is a doctoral student in the Curry School of Education and Frank Batten School of Leadership and Public Policy at the University of Virginia.

**Thomas R. Bailey** is the George and Abby O’Neill Professor of Economics and Education, Director of the Community College Research Center, and Director of the Center for Analysis for Postsecondary Education and Employment at Teachers College, Columbia University.
Bruce D. Baker is Professor in the Department of Educational Theory, Policy and Administration, Graduate School of Education at Rutgers University.

Daphna Bassok is Assistant Professor of Education and Public Policy at the University of Virginia.

Clive R. Belfield is Associate Professor of Economics at Queens College, City University of New York and Co-Director of the Center for Benefit-Cost Studies in Education at Teachers College, Columbia University.

Eric P. Bettinger is Associate Professor in the Graduate School of Education at Stanford University.

Robert Bifulco is Associate Professor of Public Administration and International Affairs in the Maxwell School of Citizenship and Public Affairs at Syracuse University.

Kevin Booker is a Senior Researcher at Mathematica Policy Research.

Brian O. Brent is the Earl B. Taylor Professor and Associate Dean of Graduate Studies, Warner Graduate School of Education and Human Development at the University of Rochester.

Dominic J. Brewer is Gail and Ira Drukier Dean of the Steinhardt School of Culture, Education, and Human Development and Professor of Education, Economics and Policy, Department of Humanities and Social Sciences in the Professions at New York University.

Katrina Bulkley is Professor of Educational Leadership at Montclair State University.

Jay G. Chambers is Senior Research Fellow and a Managing Director at the American Institutes for Research.

Sean P. Corcoran is Associate Professor of Educational Economics, New York University Steinhardt School of Culture, Education, and Human Development.

Joshua M. Cowen is Associate Professor of Education Policy in the Department of Teacher Education, College of Education at Michigan State University.

Morgaen L. Donaldson is Associate Professor of Educational Leadership at the University of Connecticut, and a Research Affiliate of the Project on the Next Generation of Teachers at Harvard University.

Thomas A. Downes is Associate Professor of Economics, Department of Economics at Tufts University.

William D. Duncombe, now deceased, was Professor of Public Administration in the Maxwell School of Citizenship and Public Affairs at Syracuse University.

Eric R. Eide is Department Chair and Professor of Economics at Brigham Young University.

William N. Evans is the Keough-Hesburgh Professor of Economics at the University of Notre Dame.

David N. Figlio is the Orrington Lunt Professor of Education and Social Policy and of Economics and Director of the Institute for Policy Research at Northwestern University.

Edward B. Fiske, former Education Editor of The New York Times, is an education writer and editor.

Patricia Gándara is Research Professor of Education and Co-Director, Civil Rights Project/Proyecto Derechos Civiles, at the University of California at Los Angeles.

Brian Gill is a Senior Fellow at Mathematica Policy Research.

Dan Goldhaber is the Director of the Center for Education Data and Research and a Research Professor in Interdisciplinary Arts and Sciences at the University of Washington Bothell.
Nora E. Gordon is Associate Professor of Public Policy at Georgetown University.

Preston C. Green is the John and Carla Klein Professor of Urban Education and a Professor of Educational Leadership and Law at the University of Connecticut.

James W. Guthrie is Professor Emeritus, Department of Leadership, Policy and Organizations, College of Education, Vanderbilt University.

Jesse Hahnel is an attorney and Director of NCYL’s FosterEd.

Michelle Hall is Dean’s Fellow, Rossier School of Education at the University of Southern California.

Jane Hannaway is Vice President and Director, National Center for Analysis of Longitudinal Data in Education Research, at the American Institutes for Research.

Janet S. Hansen, now retired, was Senior Policy Researcher at RAND Corporation.

Tenice Hardaway is Dean’s Fellow, Rossier School of Education at the University of Southern California.

Jenifer J. Harr-Robins is a Principal Researcher at the American Institutes for Research.

Douglas N. Harris is Associate Professor of Economics and University Endowed Chair in Public Education at Tulane University, and Director of the Education Research Alliance for New Orleans.

Guilbert C. Hentschke is the Richard T. Cooper and Mary Catherine Cooper Chair in Public School Administration at the University of Southern California’s Rossier School of Education.

Stephen P. Heyneman is Professor of Public Policy and International Education, Peabody College at Vanderbilt University.

Eric A. Houck is Associate Professor in Educational Leadership and Policy at the University of North Carolina at Chapel Hill.

Rebecca Jacobsen is Associate Professor of Teacher Education at Michigan State University’s College of Education.

Melinda Mechur Karp is a Senior Research Associate at the Community College Research Center at Teachers College, Columbia University.

Kieran M. Killeen is Associate Professor, College of Education and Social Services at the University of Vermont.

William S. Koski is the Eric and Nancy Wright Professor of Clinical Education and Professor of Law at Stanford Law School.

Bommi Lee is a doctoral student at Vanderbilt University.

Henry M. Levin is the William Heard Kilpatrick Professor of Economics and Education and Director of the National Center for the Study of Privatization (NCSPE) at Teachers College, Columbia University.

Susanna Loeb is the Barnett Family Professor of Education at Stanford University, Director of the Center for Education Policy Analysis, and Co-Director of Policy Analysis for California Education.

Patrick J. McEwan is Professor of Economics at Wellesley College.

Therese J. McGuire is Professor of Strategy in the Kellogg School of Management at Northwestern University.
Phuong Nguyen-Hoang is Assistant Professor, School of Urban and Regional Planning and Public Policy Center at the University of Iowa.

Allan R. Odden is Professor Emeritus of Educational Leadership and Policy Analysis at the University of Wisconsin-Madison.

John P. Papay is Assistant Professor of Education and Economics at Brown University.

Leslie E. Papke is Professor of Economics and Director of Graduate Studies at Michigan State University.

Tom Parrish is a Managing Director at the American Institutes for Research.

Lawrence O. Picus is Professor of Education Finance and Policy, Rossier School of Education at the University of Southern California.

Sean F. Reardon is Professor of Education and Director, Stanford Interdisciplinary Doctoral Training Program in Quantitative Education Policy Analysis, at Stanford University.

Andrew Reschovsky is Fellow at the Lincoln Institute of Land Policy and a Professor Emeritus of Public Affairs and Applied Economics at the University of Wisconsin-Madison.

Jennifer King Rice is Professor of Education Policy and Associate Dean for Graduate Studies and Faculty Affairs, College of Education at the University of Maryland.

Joseph P. Robinson-Cimpian is Assistant Professor, Educational Psychology at the University of Illinois at Urbana-Champaign.

Richard Rothstein is a Research Associate of the Economic Policy Institute and Senior Fellow of the Chief Justice Earl Warren Institute on Law and Social Policy at the University of California (Berkeley) School of Law.

Russell W. Rumberger is Professor of Education, Gervitz Graduate School of Education at the University of California, Santa Barbara.

Kai A. Schafft is Associate Professor of Education and Director, Center on Rural Education and Communities, at The Pennsylvania State University.

Amy Ellen Schwartz is the Daniel Patrick Moynihan Professor of Public Affairs at Syracuse University, and Director of the Institute for Education and Social Policy at New York University.

John W. Sipple is Associate Professor, Department of Development Sociology at Cornell University.

Matthew G. Springer is Assistant Professor at Peabody College Vanderbilt University and directs the National Center on Performance Incentives.

Leanna Stiefel is Professor of Economics and Education Policy at the Wagner Graduate School and the Steinhardt School, New York University.

Tien Le is Dean’s Fellow, Rossier School of Education at the University of Southern California.

Eugenia F. Toma is the Wendell H. Ford Professor of Public Policy in the Martin School of Public Policy at the University of Kentucky.

Ericka S. Weathers is a doctoral student at Stanford University.

Kenneth K. Wong is the Walter and Leonore Annenberg Chair for Education Policy at Brown University.

James Wyckoff is Curry Memorial Professor of Education and Policy and Director of the Center on Education

567
Policy and Workforce Competitiveness at the Curry School of Education at the University of Virginia.

**John Yinger** is Trustee Professor of Public Administration and Economics in the Maxwell School of Citizenship and Public Affairs at Syracuse University.

**Ron Zimmer** is Associate Professor of Public Policy and Education in the Peabody College at Vanderbilt University.
Index

Page numbers in *italics* indicate figures and tables. Page numbers followed by "n" indicate endnotes.

A

Aaronson, D. 402
Abbott v. Burke 288–2
Abdulkadiroglu, A. 47, 435
Abecedarian project 317–18
ability standard 233
absenteeism 626
absorption capability 632
Abu-Ghaida, D. 626
academic achievement see student achievement
Accelerated Learning Program (ALP) 647
Accelerated Study in Associate Programs (ASAP) 649
Accelerated Learning Program (ALP) 647–8
access: Chile 654–5; Cleveland 446; Colombia 453; District of Columbia 449; Education for All 299–303; India 455–6; Milwaukee 448; vouchers 653–4
accountability: agency theory and 34–5; balanced 83–4; disadvantaged students 551–2; as educational goal 83–4; Efficiency and 13, 14; market-based 551–7; "new" 49; performance-based 73; philanthropy and 447; productivity in education research 135, 136; special education 579; standards-based 135, 136; test-based 551–7; transient students 633–4; see also school accountability
achievement gaps: about 493; Asian-white 497, 499; between-school versus within-school patterns 592–3; causes 513–15; changes as children progress through school 496–500; changes over time 493–4; future research 505–6; growth across achievement range 509–10; Hispanic-white 495, 496, 497, 500, 502; income and 495, 497, 600; labor market effects of 509–10, 507; Mexican-white 494, 495; parental education level and 493, 494, 495–6, 497, 499–500; socioeconomic 493, 494, 495–6, 497, 499–500; socioeconomic status and 501–2; see also black-white achievement gaps: early childhood
adequacy: application of 257n10; calculations, updating 254; defined 244; educational finance reform litigation 47–9, 48–9, 231; educational goals 88; evidence-based approach 259–1; finance equity synthesis framework 230, 240; future research 254–6; history 12–13, 18n1; intergovernmental aid formulas and case studies 280–1, 290–2; measures, using 253–4; measuring 247–53; productivity in education and 137–5; professional judgment/resource cost model approach 248–9, 255; property tax 381–2, 389n11; quest for 31–13; rulings based on 362, 361–3; school district cost-function approach 251–3; special education 579–80; state litigation 47–9; successful-districts approach 242–5, 255–6; vertical equity and 236–8; vertical equity as 239, 239–40
ADM (average daily membership) 632
administrative accountability see school accountability
administrative rents 409
adult outcomes 512
adverse selection problems 34–5
AEPF (Association for Education Finance and Policy) 21, 36n1
AERA (American Education Research Association) 36n1
Africa 307–8, 309–10
AFT (American Federation of Teachers) 21
agency theory 32, 34–5, 185
aging of population 384–5
allocative Efficiency 141
ALP (Accelerated Learning Program) 647
Alternative Minimum Tax (AMT) 385, 389
alumni clubs 339
American Education Research Association (AERA) 36n1
American Federation of Teachers (AFT) 71
AMT (Alternative Minimum Tax) 385, 389
Andhra Pradesh School Choice Experiment 455–6
Angrist, J. 129, 435–6, 453
Annenberg Challenge 343
annual yearly progress (AYP) 575
APPAM (Association for Policy Analysis and Management) 21, 36n1
APPE (average per-pupil expenditure) 647
Arizona 59, 57–28, 597, 598
Arkansas 283–3, 290–1, 289n5–6, 646–17
ASAP (Accelerated Study in Associate Programs) 447–8
Asian-white achievement gaps 597; early college readiness 647; linguistic minority students 595–6; national 111–12; regional 112–13; teacher effectiveness 590–1; see also international large-scale assessments (ILSAs)
Association for Education Finance and Policy (AEFP) 21, 36n1
Association for Policy Analysis and Management (APPAM) 21, 36n1
Association of Baltimore Area Grantmakers 346
at-risk students 271–7
attachment, school 615
Atwill, K. 613
audit culture 115
Augenblick and Myers, Inc. 249
average daily membership (ADM) 432
average per-pupil expenditure (APPE) 575
AYP (annual yearly progress) 329

B

Baldwin Park (California) school district 56n9
Baltimore 549
Barley, Z. 613
Baumgartner, F. R. 66–7
Baumol’s cost disease 131
Bayer, P. 218
BEA (Bilingual Education Act) 523, 589
Beesley, A. 613
Belfield, Clive 448–9
Benavot, A. 114, 115
benefit-cost ratio 151
benefits minus costs metric 151
benefit standard 233
benefit tax 382–3, 384
Berne, Robert 11, 231–3, 245–6
Berne-Stiefel framework 231–3
Bettinger, E. 451, 453
Beverly Hills (California) school district 56n9
bias concerns 179, 181, 266–9, 366–7

570
census-based systems, for special education funding 573–4
center-based early childhood programs 517–21
Center for Research on Education Outcomes (CREDO) 434
Center of Special Education Finance (CSEF) 568–9, 573
Center on Education Policy 83
centralization of school finance 360–1
CEO America (Children’s Educational Opportunity Foundation) 450
certification, alternative 169–70, 526, 613–14
CFA (cost function analysis) 309–10, 277, 557–8
Chambers, J. 497–8, 509–10, 600, 613
Channel One 350n32
Charlottesville Summit 328
charter management organizations (CMOs) 426–7, 475–6
charter schools: about 16, 422–3, 470–1; academic achievement 432–3; civic socialization and 222; competitive effects of 413, 274, 118; current landscape 473, 475, 475–6; disadvantaged students 557–9; disadvantaged students, isolation of 430–1; educational resource allocation 20–1, 149–50; financing 473, 475; instructional conditions 428; intergroup exposure 429–30; KIPP 436, 437; lotteries for 277, 434–6; “No Excuses” model 427; organizational structures 426–8; policy design and implementation 424–6; Portfolio Management Model 425; rationales for 446; by state 470–1, 472; student fixed effects studies 433–4; student sorting and school competition 219–20; supply and demand for 476–7; teacher qualifications 428–9; see also privatization, educational cheating by teachers 201
Chicago: accountability, school 202; Child-Parent Centers 153–4, 154, 518–19, 521; Gautreaux program 547–9; political context of education finance 71, 74; promotion/graduation exams 554; Teacher Advancement Program 592; teacher evaluation 187–8
Child-Parent Centers 153–4, 524, 518–19, 521
Children’s Educational Opportunity Foundation (CEO America) 450
Children’s Scholarship Fund (CSF) 450–1, 457
Chile 217, 454–5, 457
China 114, 115, 118n2, 306–7
Chingos, Matthew 451
Cincinnati 176, 178, 183, 187
Citizenship Clause 33
City University of New York 647–8
civic education 29, 223–3
Civil Rights Act (1964) 322, 324
civil rights 322–3, 328–7
Civil Rights Act 450
classroom observations, standards-based 178–80
class size 89, 129–30, 593
Cleveland 220, 448–9, 456, 464
Clinton, Bill 72–3
Clotfelter, C. T. 500
clubs, school-based parent, alumni, and booster 339
CMOs (charter management organizations) 426–7, 475–6
Coase, Ronald 34
Cobb-Douglas function 274n6
coefficient of variation, in per-pupil spending inequality trends 358
cognitive skills 504
Cold War 322
Coleman, James S. 13
"Coleman Report" 13, 235
COL (cost-of-living) index 284–5
college 555, 608–9; see also high school-college divide
Colombia 453–3, 457

572
Colorado 464
Colvin, R. L. 342
combination formulas 285–6
commercial partnerships 348–9
Commission on the Reorganization of Secondary Education 79
Common Core State Standards (CCSS) 344–5, 396, 641–2
common-school model 221
Community College of Baltimore County 647
community foundations 340
comparison group 99
competition see school competition
competitive positional good, education as 237–8
competitive wage index (CWI) 265–6
Comprehensive School Improvement Program (CSIP) 343
compulsory attendance laws 480
corporate supporters 347–9
cost-benefit analysis (CBA): costs analysis 143–6, 146; discount rate 151, 155; Efficiency and 141; empirical and methodological challenges 152–3; examples and findings 153–4, 255; metrics 151–2; opportunity cost 147–8; origins 142–3; shadow pricing techniques 150–1, 152–3; use 142, 155
cost differentials: about 260–1; disadvantaged students 270–3; education cost functions 261–3; enrollment size 689–70; futures research 273–4; geographic variation in resource prices 264–8
cost dimension, in school finance formulas 286–7
cost-effectiveness analysis (CEA): about 146; cost-effectiveness ratios 146–7; costs analysis 143–6, 148; Efficiency and 141; empirical and methodological challenges 147–8; examples and findings 148–50, 249; opportunity cost and 141–2; origins 142–3; use 142, 155
cost-effectiveness ratios 146–7
cost function analysis (CFA) 269–70, 272, 597, 607–8
costing-out 45–9, 577–80
cost-of-living (COL) index 264–5
costs analysis 143–6, 148
cost studies, for linguistic minority students 596–9, 597–8, 600, 601
cost-utility analysis (CUA) 146
counterfactuals 89–90
Crain, R. L. 547
CREDO (Center for Research on Education Outcomes) 434
Crowley, M. 608
CSEF (Center of Special Education Finance) 544–9, 573
CSF (Children’s Scholarship Fund) 450–1, 457
CSIP (Comprehensive School Improvement Program) 343
Cubberley, Elwood P. 3
Curriculum 113, 609–10, 616
Currie, J. 531–2
CWI (competitive wage index) 265–6
cyber schools 427–8, 482–3

D

Dakar Framework for Action 298, 300, 301, 303
Danielon, Charlotte 176, 178
DD (difference-in-differences) 97–8, 99
Dee, T. 98, 188, 536, 539, 556
demographics 45, 384–5
Denver 183
Department of Defense schools 528–9
descriptive research 87, 88
desegregation: black achievement and 546–7; blacks, long-term effects on 551–2; federal role in education finance 322–3; housing opportunity experiments 547–8; income-based 551–2; Milwaukee 446; peer effects and achievement 548–51; school competition and 219–21
developmental bilingual programs 597, 599
developmental education 645–8
difference-in-differences (DD) 97–8, 99
disabled students 200, 217, 271–2, 326–7; see also special education
disadvantaged students: about 545–6, 559–60; accountability, market-based 557–8; accountability, test-based 553–7; cost differentials 270–3;
desegregation 546–51; isolation of 449–50; resources and funding 552–3
discontinuity assignment 94–5, 99–10, 108n15, 108n16
discount rate 151, 155
discrepancy model 578
disequilibrium 66–7
distance education 609–10
District of Columbia: charter schools and civic socialization 223; evaluation and incentive system 167, 184, 187, 188, 535, 539; vouchers 449, 457, 465
districts see school districts
diverse objectives problems 35
donor-advised funds 346
Donors Choose 346
Douglas County, Colorado 464
Downes, T. A. 398–9, 401, 403
Downey, Lawrence 80–1
Dronkers, J. 223
dropout rates 366, 555
dual enrollment 642–5
dual-immersion programs 592, 597
dubner, Stephen J. 23
Dumaret, A.-C. 513–14
Duncan, G. J. 548
Duncombe, W. 251–2, 255, 597, 607–8
Dwyer, M. 513–14
dye, R. F. 398, 399, 481–2
Dyarnski, S. M. 98

E

574
Early Childhood Longitudinal Study (ECLS) 519
Early Childhood Longitudinal Study-Kindergarten (ECLS-K) 496, 497, 507
Early College High Schools (ECHS) 644
early college readiness assessments 642
early-exit programs 591–3, 592
earnings 27–5, 28, 364; education services, allocating 27–28, 32; human-capital theory 24–5
Easton, David C. 61
ECHS (Early College High Schools) 644
ECLS (Early Childhood Longitudinal Study) 519
ECLS-K (Early Childhood Longitudinal Study-Kindergarten) 496, 497, 507
ecology 66
economic development 299, 303
ECONOMICS in education policy research: about 21–3; economics, overview of 23–4; educational organizations as economic enterprises 23–5; education and earnings 27–28, 26, 364; education services, allocating 27–28, 32; human-capital theory 24–5
ECONOMICS of size in education 268–9, 617
EDI (Education for All Development Index) 302
educational finance reform litigation: adequacy 48–9, 281; equity 280; evolution of 45–5; federal equal protection litigation 45–5; future of 52–5; intellectual and legal roots 41–2; issues, current 48–51; “new accountability” 49; special education 579–8; standards-based reform 48–9; state “adequacy” litigation 47–8; state “equity” litigation 46; state role in education finance 355, 356, 357, 372–5; system-wide reform 48–51, 57n30; see also specific cases
educational goals: about 77; current consensus on 82–4; linguistic minority students 588–9; pre-twentieth century 78–9; synthesizing 89–3, 80
educational governance 86–7
educational organizations as economic enterprises 52–4
educational practices studies 136–7
educational privatization see privatization, educational
educational service agencies (ESAs) 617–18
education clauses of state constitutions 380
education cost functions 261–3
education finance and policy, development of 8–31
Education for All (EFA) 497–303
Education for All Development Index (EDI) 302
Education Law Center 247
education management organizations (EMOs) 426–7, 471, 475
education policies 108–13, 110
Education Policy Improvement Center 255
education production function 19–4, 147–34
education services, allocating 27–27, 32
education tax credits (ETCs) 445
Education Week 246–7
EEOA (Equal Educational Opportunities Act) 247
EFA (Education for All) 497–303
effective equity 43–4
effectiveness, teacher see teacher effectiveness
efficiency: allocative 143; cost-benefit analysis 143; cost-effectiveness analysis 143; educational finance reform litigation 23–4; political context of education finance 67, 67–2; privatization, educational 418; productive 415; productivity in education 147; property tax 382–3; quest for 23–15; school 199–200; school district 262–3; special education 586; technical 143
Eide, E. R. 32

grants 6–7; linguistic minority students, educational goals for 588–90; political context of education finance 69–70; special education funding 574–5; see also federal government role in school finance: specific programs and laws
federal government role in school finance: about 317–19; evolution in United States 319–28, 320; future options 331–2; intergovernmental aid formulas and case studies 279; in twenty-first century 328–31; see also federal government
federal range ratio 245–6
fees, user 305–11
FEP see Fluent English Proficient (FEP)
Feuer, M. J. 117
Figlio, D. N. 80, 217, 298–9, 401, 403
FIMS (First International Mathematics Study) 106
finance equity synthesis framework 238–40, 239
financial incentives for teachers 532, 535, 539, 612–13
financing, as government role 29
Finland 115, 117
First International Mathematics Study (FIMS) 106
fiscal effects of education finance reform: adequacy-based rulings 360–1; court-mandated reform versus legislative reform 364–5; equity-based rulings 363–4; evidence across states and over time 361–5; evidence from specific states 360–1; level of spending 367; tax and expenditure limits 394–7
fiscal neutrality 44–5, 234–5, 239, 239
fiscal resources 597–8, 597–9
fiscal stress 501(c)(3) organizations 338
flat grants 8
Florida 280, 315, 417, 477
Fluent English Proficient (FEP) 586, 587; see also linguistic minority students
Ford Foundation 343
Foundation Center 338–9
foundation philanthropy 341–7
foundation programs 9, 202–3, 207, 223–4
Founding Fathers 78
Fourteenth Amendment 42, 43–5, 53, 233–4
Fowler, W. J. 265
Framework for Teaching (Danielson) 176, 178
Franklin, Benjamin 78
Freakonomics (Levitt and Dubner) 43
free and reduced price lunch 283, 295n5
free appropriate public education (FAPE) 326–7, 568
Friedman, Milton 15, 30, 36n5, 444
Fryer, R. G. 501, 503, 507nn11–12, 514
full funding for special education 575
fundamentalist Christian schools 222–3
"fundamental right" test 42

G

Galindo, C. 488–9
Gallego, F. A. 455
GAO (Government Accountability Office) 344–5
Garms, W. I. 242n18
Gates Foundation 344
Gautreaux program 547–8
GCEI (geographic cost of education indices) 264–5
gender equity 303–5
General Education Board (GEB) 342
genetic factors, in achievement gaps 513, 514–15
geographic cost of education indices (GCEI) 264–8
GER (gross enrollment ratio) 300–1
Germany 109–10
GI Bills 322
Gingrich, Newt 24, 75n4
Gini coefficient 256n4, 358, 359, 359–60
Glass, Gene 149
goals, educational see educational goals
Godwin, J. R. 222–3
Godwin, R. K. 222–3
Goldhaber, Dan 12, 165
Goodlad, John 81
Gordon, N. E. 325–6
Gormley, W. T. 520
governance, educational 24, 4–7
Government Accountability Office (GAO) 348–9
government functions 29
government structures 69–72
graduation rates, high school 149, 149–50
Grant, James 298
Grantmakers for Education 345
Great Depression 79–80
Greene, J. P. 217, 344, 554–5, 560n1
Grimshaw, W. 71
Grodzins, Morton 69
Gross Domestic Product percentage 4, 4, 27, 27
gross enrollment ratio (GER) 300–1
growth-model approach 198, 199, 200, 203
guaranteed tax base (GTB) programs 283–4, 294–5
Guthrie, J. W. 15–16

H

Haig, Robert M. 9
Hammer, P. C. 613
Harris, D. N. 556
Harris, William T. 13
Hart, B. 511
Hart, C.M.D. 217
Hartford, Connecticut 51
Hartman, W. T. 576
Hawaii 232, 241n1
Head Start 227–8, 318, 319, 321–2
Heckman, J. 512–13
hedonic teacher cost index 266–7
Herrington, C. D. 556
Hess, F. 203, 346–7
high cost risk pools 576
"highly qualified teachers" 167n4, 611–12
high politics 65–6
high school-college divide: about 640–1; developmental education 645–8; dual enrollment 642–5; future research 649–50; high school-college alignment 641–2
high school dropout rates 366, 555
high school graduation rates 149, 149–50
High/Scope Perry Preschool experiment 512, 517, 518
Hispanic-white achievement gaps 493, 494, 495, 497, 498–9, 502
Hispanic-white wage gap 504
Hollands, F. 149, 149–50
Holley, M. J. 616–17
homeless students 628
homeostasis 63–4, 65
homeschool charters 627–9, 682–3
home schooling 470–1, 472, 480–2
home visitation programs 516–17
Hoople, D. S. 32
horizontal equity: defined 232, 245; educational finance reform litigation 43; fiscal neutrality and 239, 239; measures 245–6; in tax policy 233; see also equity
horizontally distributed power 70–1
Horne v. Flores (2009) 590
Horowitz, Harold 43, 44
housing opportunity experiments 547–8
Howley, A. A. 614
Howley, C. B. 614
Hoxby, C. M. 96, 218, 318, 345, 435, 476, 549–50
Hsieh, C. 454, 455
hukou system 114, 118n2
human-capital theory 24–5, 504
human rights 299, 303

I
I-BEST program 647
ICER (incremental cost-effectiveness ratio) 147
IDEA see Individuals with Disabilities Education Act (IDEA)
IEP (individualized education program) 568, 579–80
Illinois 71, 401–2, 405n8
ILSAs see international large-scale assessments (ILSAs)
Impact Aid program 321–2
IMPACT system 167, 184, 187, 188, 535, 539
Imazeki, J. 340
incentive problem 35, 253–4
income: desegregation based on 551–2; language exposure differences and 511; socioeconomic achievement gaps 493, 497, 500; tax and expenditure limits and 399
income tax 386–7
incremental cost-effectiveness ratio (ICER) 147
independent foundations 341–7
India 455–6, 457
Indiana 31, 78, 465
individualized education program (IEP) 568, 579–80
Individuals with Disabilities Education Act (IDEA): about 567–8; equity and adequacy in school finance 236, 242n17; federal role in education finance 326; funding for special education 574–5; special education encroachment 578
induced costs 146
industrialization 7–8
inefficiency 131–2
inequality see equality
information, incomplete 28–2
information asymmetry problems 35
Ingersoll, R. 611
ingredients method of costs analysis 143–5, 145
input-output research 134–5
inputs, school 126, 128–31, 138n1
inspection systems 195, 206–7
institutional perspective 135–6, 149
instrumental variables 99–7, 99
interactive television (ITV) 610
intergovernmental aid formulas and case studies: about 279–81; adequacy 280–1, 290–2; equity 280, 287–9; school finance formulas 281–7, 293, 293–5; transient student populations 632–3
intergroup exposure 429–30
internal rate of return (IRR) 151–2
internal validity 89
International Association for the Evaluation of Educational Achievement 106
interstate inequality 53
intradistrict funding inequality 52–3
iron triangle of politics 64–5
IRR (internal rate of return) 151–2
Israel 129
ITV (interactive television) 610

J
Jacob, B. A. 554, 556
Japan 115
Jefferson, Thomas 78
Jencks, C. 548–9
Johnson, Lyndon Baines 66
Johnson, S. M. 185
Jones, B. D. 66–7
Jomtien Declaration 298, 299, 301

K
Kain, J. F. 549–50
Kang, J. 435
Kansas 249
Karoly, L. A. 154, 154, 155
Kentucky 285–6, 287–9
Kentucky Supreme Court 12, 47
Kenya 307–8
King, F. 453
King, J. K. 130
Kingdon, J. W. 66
KIPP charter schools 436, 437
Kirp, David 41–4
Klasen, S. 304
Koski, W. S. 237–9
Kremer, M. 453
Krueger, A. B. 93, 129, 138
Kugler, M. 453
Kyrgyzstan 110–11

L

labor market effects of achievement gaps 503–5
labor unions 71, 78, 177–8, 400, 401
Ladd, H. F. 500, 548
land grants 5–7
language exposure differences 511
language minority students see linguistic minority students
Lanham Act (1940) 321–2
late-exit programs 592, 592
Latin America 112–13
Lavy, V. 189
LCFF (Local Control Funding Formula) 291
learning communities 646
least restrictive environment (LRE) 568
Lee, V. E. 521
LEFs (local education funds) 339–40
legal culture 68
legislative culture 67
leveling up versus leveling down effect 360
Levin, Henry 149
Levine, J. 98
Levitt, S. D. 23, 501, 503, 607
m11–12, 514
Liberty 13–17, 62–63, see also school choice
licensure status 152–60
Liebman, James 55
limited English proficiency see English learners (ELs)
linguistic minority population 585–8, 596
linguistic minority students: about 585; civil rights for 223–4; classifications 586; cost studies 586–601; educational goals 588–91; educational programs 591–3, 592; instructional materials 592; instructional programs 591–2, 592; instructional time 591–2; linguistic minority population 587–8, 586; material resources 592–900, 600; 601; poor and low-income students compared to 601–2; 602; resource needs, additional 601–3, 602; resource requirements of educational programs 593–6; school performance 566–7, 583; school readiness 587, 587–8
literacy practices, early 516
litigation, educational finance reform see educational finance reform litigation
Liu, Goodwin 53
LM students see linguistic minority students
lobbying groups 72
Local Control Funding Formula (LCFF) 291
local education funds (LEFs) 339–40
local funding of schools see property tax
local governments: education governance role 7; education revenue 6, 8, 320, 356; political context of education finance 69, 70
local school districts see school districts
Loeb, S. 35, 521
Long-Term Trend NAEP (NAEP-LTT) 491, 492, 493
Los Angeles Unified School District 52, 164
lotteries for charter schools 97, 434–6
Louisiana 31, 465, 478
low-performance schools see teacher effectiveness
LRE (least restrictive environment) 568
Ludwig, J. 96, 519, 548
lunch, free and reduced price 283, 295n5

M

McDonnell, L. 67
McEwan, P. J. 217
McGuire, T. J. 386, 398, 409, 401–2
McKay Scholarship Program 417, 477
McKinney-Vento Homeless Assistance Act (2002) 628
McLoone index 246
McMillan, R. 218
McMillen, D. P. 402
McNeill, L. M. 348
macro policy change, preconditions for 66–7
Mader, N. S. 215
magnet schools 16, 469
Magnuson, K. A. 520–1
Mahard, F. 547
Maier, A. 611
Maine 466
Main NAEP 491–2, 494
Mann, Horace 78–9
market failure 28–9
markets 28, 68–2
Martinez-Ebers, V. 222–3
Massachusetts: charter schools 435–6; educational goals 78–8; lotteries for charter school enrollment 97; per-pupil state revenues 98; political context of education finance 61–2
maternal education 515–16
mathematical treatment of education production function 132–4
Mathematica Policy Research 436
mathematics: accountability 206; charter schools 436; developmental education 647; tax and expenditure limits 399; Teach for America teachers 614; see also achievement gaps
Mayer, C. J. 399
Mayer, S. E. 548–9
measurement error 199

582
Measurement of Equity in School Finance, The (Berne and Stiefel) 11
Medina, C. 453
Meister, Gail 149
MEP (Migrant Education Program) 627
merit pay 14–15, 166, 168n21
Merrill, L. 611
METCO program 546–7, 551
Mexican-white achievement gaps 494, 495; see also Hispanic-white achievement gaps
Meyer, H.-D. 114, 115
Miami-Dade County Public Schools 536–7
Michigan 379
micro politics 64–5
migrant and immigrant students 626–7
Migrant Education Program (MEP) 627
military families, mobile students from 628–9
Mill, John Stuart 24, 25
Millennium Development Goals 299
Miller, D. L. 519
Milwaukee 17, 31, 98, 215, 447–8, 456, 466
Minnesota 633
Missouri 285
mobile students see transient student populations
Monk, D. H. 130
Montana 249
Monte Carlo simulations 153
Montgomery County, Maryland 176, 341
Montoy v. Kansas (2003) 422n12
Morrill Land Grant Acts 332n4
Moving to Opportunity (MTO) program 547, 548, 551
Mullins, D. R. 293–6
Murkaka, S. 435
Murray, S. E. 11, 33–4
Musgrave, Richard 233

N

NAEP (National Assessment of Educational Progress) 118n1, 201, 202, 204–5, 491–2, 493, 494
NAEP-LTT (Long-Term Trend NAEP) 491, 492, 493
National Academy of Education 81
National Assessment of Educational Progress (NAEP) 118n1, 201, 202, 204–5, 491–2, 493, 494
national assessments 111–12
National Board for Professional Teaching Standards (NBPTS) 161, 165, 176
National Center for Education Statistics (NCES) 81, 377
National Commission on Excellence in Education 8, 72, 73, 328
National Conference of State Legislatures (NCSL) 397, 598
National Defense Education Act (1958) 322
National Education Association (NEA) 71, 72–80
National Public Education Financial Survey (National Center for Education Statistics) 377
National Research Council 87
National Vocational Education Act (1917) 320
Nation at Risk, A (National Commission on Excellence in Education) 8, 72, 73, 328

583
NBPTS (National Board for Professional Teaching Standards) 161–165, 176
NCES (National Center for Education Statistics) 81, 377
NCLB see No Child Left Behind Act (NCLB)
NCSL (National Conference of State Legislatures) 207, 298
NEA (National Education Association) 21, 29–40
Nebraska 249
Nechyba, T. S. 550
needs-based funding equality 42
neighborhood attendance zones 529
net enrollment ratio (NER) 131
net present value (NPV) 151
"new accountability" 49
New Haven, Connecticut 187
New Jersey 46, 80, 284–9
NewSchools Venture Fund 345–6
New York City 435, 558
New York State: Children’s Scholarship Fund 450–1; linguistic minority students 597, 597–9; racial isolation 141; school district cost-function approach 251–2; special education funding 376
New Zealand 220
NGOs see nongovernmental organizations (NGOs)
95-to-5 ratio 358
Nita, K. A. 616–17
No Child Left Behind Act (NCLB): about 241; accountability 14, 23–24, 283–4, 526–7; agency theory and accountability 527; annual yearly progress 329; beginnings 22–3; challenges to accountability 24–5; disadvantaged students 555–6; educational goals 23; equity and adequacy in school finance 236; federal government role in school finance 328–30; “highly qualified teachers” 162–4; 611–12; linguistic minority students 589–90; political context of education finance 72–4; political dynamics, post-NCLB 73–4; waiver process 329
"No Excuses" model 427
noncensus systems, for special education funding 274
nonclassroom-based charters 422–3, 482–3
noncognitive skills 505, 522
nonexperimental research 90–91
nongovernmental organizations (NGOs): about 336–7; corporate supporters 347–8; data limitations 338–9; equity concerns 340–1; locally-based voluntary contributors 399–41; political context of education finance 72–4; private independent foundations 341–7; sustainability concerns 341; types 327–8
North Dakota 283
Northwest Ordinance (1787) 78
NPV (net present value) 151
Nursing Home Visitation Program 516–17

O
Oates, W. E. 386, 387
Obama, Barack 75, 330
O’Day, J. 12
OECD (Organisation for Economic Co-operation and Development) 106, 108, 115
Ohio 55, 466, 477, 478
Oklahoma 523
OLS (ordinary least-squares regression) 100n6
"one scholar, one dollar" principle 43
on-the-job learning 538
open-enrollment schools 469

584
operation, as government role 29
opportunity cost 241–2
Opportunity Scholarship Program 215, 217, 477
opportunity surplus 238
optimization 24
ordinary least-squares regression (OLS) 100–6
Oregon 480, 576–7
Organisation for Economic Co-operation and Development (OECD) 106, 108, 115
outcome equity 235–6
outcomes: adult 512–6; education production function 127–8; employment 555; equal outcomes standard 43–5; 239, 212
privatization, educational 414; special education 580–1
outputs, school 125–6, 127–8

P

PACES voucher program 452–3, 457
Papay, J. P. 185
PAR (Peer Assistance and Review) 185, 186
parents: clubs for 339; educational goals 82–3; education level 493, 494, 495–6, 497, 499–500; engagement of 615; school accountability and 196–7; surveys of 183–4
Parents Involved in Community Schools v. Seattle School District #1 (2007) 323
Parrish, T. 576
Partnership for Educational Revitalization in the Americas (PREAL) 112–13
PASEC (Program d’analyse des systems educatifs des pays dela CONFEMEN) 112
Pathak, P. A. 435–6
Peer Assistance and Review (PAR) 185, 186
peer effects 548–51
Pennsylvania 78, 452–1
percentage equalizing plans 2
performance-based accountability 71
performance evaluation see teacher evaluation
performance objectives 262
per-pupil spending 4, 43, 43, 257–60, 358, 359
Perry Preschool experiment 505, 512, 517, 518
Peterson, Paul E. 203, 217, 453
PGEs (promotion and graduation exams) 555
pipelines 347
Place Called School, A (Goodlad) 81
PMM (Portfolio Management Model) 415
political context of education finance: government structures 68–72; importance of 68–9; No Child Left Behind Act as example of political dynamics 72–5; political cultures 67–2; public values and political processes 67, 67–2, 64
political cultures 67–2
Portfolio Management Model (PMM) 415
portfolios, for teacher evaluation 184
Posso, C. 453
poverty 271, 299, 601–2, 602, 629–30
Powell, Walter 34
power 69–71
PREAL (Partnership for Educational Revitalization in the Americas) 112–13
preschool 517–22
pre-service teacher preparation 537–8
principal-agent theory 32, 34–5, 195
private independent foundations 341–7
private religious schools 222–3
Private Wealth and Public Education (Coons, Clune, and Sugarman) 10–11
privatization, educational: about 411–12; controversy over 412; defining 412–14; design provisions 416–20; equity 415; freedom of choice 415; funding 413; governance 413; issues in 414–20; outcomes 418; production 413–14; productive Efficiency 415; social cohesion 415–16; sponsorship 412–13; types 411–12; see also charter schools; vouchers
process research 38
production 413–14
productive Efficiency 415
productivity in education: about 33, 125–2; educational practices studies 136–7; education production function 127–34, 138n1; input-output research 134–5; institutional perspective 135–6; research characteristics 134–7; social values and 137–8; student characteristics and 126, 131
professional culture 68
professional development for teachers 166, 538, 612
professional-judgment panels: enrollment size and education costs 270; linguistic minority students 296, 307–8, 398–9, 600; measuring school finance adequacy 249–9, 255; student disadvantage and education costs 272
proficiency standards 203
Program d’analyse des systems educatifs des pays dela CONFEMEN (PASEC) 132
programmatic mandates 50
promotion and graduation exams (PGEs) 555
propensity score methods 92
property tax: about 276; administration 384; aging of population and 384–5; alternatives 385–7; alternatives to 385–7; as benefit tax 382–3, 384; court challenges and legislative actions 380–2, 389n11; demographic trends and 384–5; economic trends and 385, 389n18; education funding role 376–80, 377, 378; Efficiency 382–3; elderly and 384–5; equity 383–4; formula 393; future in school funding 387–8; income tax as alternative to 386–7; sales tax as alternative to 386, 387; stability of 387; structural concerns 382–4; are also tax and expenditure limits (TELs)
Proposition One 10–11, 43–2
Proposition 1 (California) 350, 366n15, 379, 399
Public Agenda 82–3
public confidence in schools 75n3, 81–4
public goods 29
public values and political processes 67, 68–7, 64
punctuated equilibrium 98–7
pupil funding weights 577–8
pupil-teacher ratio 4, 7
pure vertical equity 230, 239

Q

quality 292–93, 299–11; see also teacher quality
quantitative research methods: causal research, importance of 88–9; choosing 88; combining methods to improve causal inference 92–9; difference-indifferences 97–8, 99; discontinuity assignment 92–4, 99–2, 100n15, 101n16; instrumental variables 92–7, 95; propensity score methods 92; randomized assignment 92–4; regression analysis 91–2, 100n6; research, defining 87–8; research term definitions 89–90; statistical controls for observed variables 90–2
quasi-experimental research 90, 94–7
race, 25, 26, 51, 219–20, 558–9; see also achievement gaps: specific racial groups
Race to the Top, 136–7
random errors, 199
randomized assignment, 22–3
range, 245
rationality, 24
“rational review” test, 42
RDD (regression-discontinuity design), 94–5, 98–9, 100n15, 101n16
Reach Out and Read (ROR) program, 516
reading, 206, 399, 405n9, 516; see also achievement gaps
Reagan, Ronald, 73
Reardon, S. F., 495, 498–9
Reber, S., 325–6
Reckhow, Sarah, 344, 345, 346–7
Reed v. California (2010), 52
reform: court-mandated versus legislative, 364–5; market-based, 135–6; standards-based, 9–33; 12, 48–9; systemic, 425; teacher evaluation, 177–8; see also educational finance reform litigation
regional assessments, 112–13
regression analysis, 91–2, 100n6
regression-discontinuity design (RDD), 94–5, 98–9, 100n15, 101n16
regulation, 29, 416–17
regulative culture, 67–8
Reich, R., 237–8
relative deprivation theory, 549
reliability, 179–80, 181–2
religious schools, private, 222–3
remediation, 445–6
Reschovsky, A. J., 519
research: causal, 87–9; defining, 87–8; descriptive, 87, 88; experimental, 90, 91–4, 98–9; input-output, 134–5; nonexperimental, 90, 94–7; process, 88; quasi-experimental, 90, 94–5; term definitions, 89–90; see also quantitative research methods
resource cost model approach, 248–9, 255
resource distribution, 8–9
response to intervention approach, 578
retention, teacher, 531–5
revenue sources, 6, 5, 320, 354, 354–5, 378, 379
Reynolds, A. J., 519
Rich Schools, Poor Schools (Wise), 10
Risley, T. R., 511
Rivkin, S. G., 549–50
Robinson v. Cahill (1973), 46, 80
Rockefeller Brothers Fund, 80
Rockoff, J. E., 364–5
Roderick, M., 554
Romano, R., 219
ROR (Reach Out and Read) program, 516
Roscigno, V. J., 608
Rothstein, J., 218–19, 318
Rouse, Cecilia E., 98, 217, 447
rural schools: about, 607; consolidation, 616–17; curriculum, 609–10, 616; defining, 607–8; distance education, 609–10; educational service agencies, 617–18; enrollment, 609, 615–18; financial incentives for teachers, 612–13; interactive television, 510; postsecondary enrollment and completion, 608–9; professional development for teachers, 612; sparsity, 614–15; staffing, 611–14; student achievement, 608; supplementary educational
services 609; teacher quality 611–12; teacher recruitment and retention 611, 612–14; teaching, alternative pathways to 613–14; transient students 639–41; transportation for students 614–15.

S

Saavedra, J. 453
Sabel, Charles 55
salaries, teacher 14–15, 166, 168n21, 520
sales tax 286, 287
Sanbonmatsu, L. 548
SAPP (state aid per pupil) 282, 283–4, 293, 294
Sartain, L. 187–8
Sass, T. R. 549–50
SAT scores 366
scarcity 24
Scarr, S. 513
Schneider, M. 272
school accountability: consequences 196–7; design matters 196–201; disadvantaged students 556–7; exclusions 200–1; future research 297; history 194–5; indicators, scope and domains of 197–8; rationale for 195–6; standardized tests 198–200, 203; student achievement and 201–7; test-based 198–200, 203, 556–7; vouchers 479; see also accountability; school accountability; teacher evaluation
school attachment 635
school choice 74, 415, 418, 468, 470; see also school competition; specific programs
school competition: about 211–13; charter schools, competitive effects of 213, 214, 215; civic socialization and 221–3; policy implications and future research 221–3; student achievement and 213–19; student sorting and 219–21; Tiebout competition and student achievement 218–19; voucher programs, competitive effects of 215, 216, 217–18, 479
school districts: characteristics, adjusting aid formulas for 287; consolidated 614–15; cost-function approach 251–3; Efficiency 262–3; financial management role 237, 241n2; intradistrict funding inequality 52–3; school finance role 279–80; size of 268–70, 617, 631–2; transient student populations 631–2
school Efficiency 199–200
school fees 305–11
school finance formulas 281–7, 283, 293, 295–7
School Finance Reform Act (New Jersey, 2008) 289
school finance reform litigation see educational finance reform litigation
school foundations 329
School Funds and Their Apportionment (Cubberley) 3
school inputs 126, 128–31, 138n1
school outputs 125–6, 127–8
school size 631–2
school vouchers see vouchers
Schult, T. P. 304
Schwab, R. M. 13, 386, 387
Schwerdt, G. 554–5, 560n1
science education 327
Scotland 321
screening hypothesis 36n4
SEEP (Special Education Expenditure Project) 260, 521, 579, 581n12
selection bias 99
sensitivity analysis 148, 153
Serrano v. Priest (1971) 10, 56n9, 234, 360, 381
Servicemen’s Readjustment Act (1940) 322
SES (socioeconomic status) 501–2, 641, 650n2
SES (supplementary educational services) 609
Shadbegian, R. J. 395
shadow pricing techniques 150–1, 152–3
Shamblen, S. 414
sheltered English programs 992, 997
SLD (specific learning disabilities) 578
Slonim, R. 451
SLOs (student learning objectives) 182–3
Smith, Adam 24–5, 241n8
Smith, M. 12
Smith, M. C. 242n18
Smith-Hughes Act (1917) 320
social cohesion 435–16, 439
Social Security Administration 98
social values 137–8
socioeconomic achievement gaps 493, 494, 495–6, 497, 499–500
socioeconomic status (SES) 501–2, 641, 650n2
soft drink contracts 349
South Africa 307, 309
southern states 195, 207n1, 325–6
South Korea 115, 116, 117–18
sparsity, and rural schools 614–15
special education: accountability 577; adequacy 579–80; Efficiency 588; encroachment 577–8; foundation programs 283, 295n6; free appropriate public education 326–7, 568; funding 573–7; historical perspective 567–8; individualized education program 568, 579–80; least restrictive environment 568; litigation 574–9; outcomes 580–1; selective assignment of students to 200–1, 202, 257n17; spending on 568–9, 570, 571, 572, 573; unique aspects of 568; vouchers 461–2, 477
Special Education Expenditure Project (SEEP) 207, 521, 579, 582n12
specific learning disabilities (SLD) 578
spending: as Gross Domestic Product percentage 4, 47, 67, 67 per-pupil 4, 43, 357–60, 358, 359; school inputs in education production function 128–9, 138n1; special education 568–9, 570, 571, 572, 573
sponsorship 412–13
Staiger, D. 164
standards-based reform 2–31, 12, 48–2
STAR experiment 129–30
state aid per pupil (SAPP) 282, 283–4, 293, 294
state governments: “adequacy” litigation 47–3; education clauses of state constitutions 380; education governance role 6, 7; education revenue 6, 8, 320, 356, 354–5, 378; equal protection clauses of state constitutions 380; “equity” litigation 46; linguistic minority students, educational goals for 591; political context of education finance 69, 70–1; special education funding 575–7; special education spending 571; see also state role in education finance
state role in education finance: about 153–4; adequacy-based rulings 362, 360–3; court-mandated reform versus legislative reform 364–5; education finance reform and level and distribution of spending 357–65; education finance reform effects on other public expenditures 365; equity-based rulings 363–4; fiscal effects of education finance reform 360–1, 362; intergovernmental aid formulas and case studies 278–80; level of spending 362, 363; nonfiscal effects of education finance reform 365–7; per-pupil spending inequality trends 357–60, 358, 359; revenue sources 354, 355–6; see also state governments
statistical controls for observed variables 90–1, 108n6
status-based systems, of school accountability 199–200, 209
Steinberg, M. 187–8
Stievel, Leanna 11, 231–5, 245–6
strategic philanthropy 345, 346
Strayer, George D. 2
Strayer-Haig plan 2

589
student achievement: black 546–7; Chile 454, 455; Cleveland 448–9; Colombia 452–3; District of Columbia 449; early childhood environment and 512; India 456; Milwaukee 447–8; nonfiscal effects of education finance reform on 365–7; peer effects and 548–51; rural schools 668; school accountability and 207–2; school competition and 213, 214, 215, 216, 217–19; tax and expenditure limits and 397–9; vouchers 450–1, 458, 479
student fixed effects studies 433–4
student learning objectives (SLOs) 182–3
student mobility see transient student populations
students: at-risk 271–2; black 430, 521–2, 546–7, 551–2; disabled 280, 281, 271–2, 326–7; homeless 628; migrant and immigrant 626–7; from military families 628–9; surveys of 313; transportation services for 614–15, 629; white 430, 521–2; see also disadvantaged students, linguistic minority students; transient student populations
student transiency see transient student populations
successful-districts approach 249–50, 253–6
Sugarman, Stephen D. 19–11, 15, 44–5, 234–5
supplemental programs for linguistic minority students 593
supplementary educational services (SES) 600
support services 417
surveys, in teacher evaluation 183–4
"suspect classification" test 42
Sweden 218
system-wide reform 49–51, 57n30

T

TABORs (Taxpayers’ Bills of Rights) 292
Talent Search program 144–5, 145, 149, 149–50
Talent Transfer Initiative (TTI) 533
Tanzania 369
TAP (Teacher Advancement Program) 535
tax and expenditure limits (TELs): about 392–2; administrative rents and 498; enrollment and 399; fiscal effects of 394–5; impact, asymmetry of 400–3; private responses to 402; student performance and 397–2; teacher characteristics and 400–1; types 293–4; union rents and 400, 401; see also property tax
Tax Credit Scholarship Program 217
tax credits/deductions 445, 470–1
taxes: alternative minimum 385, 389n18; benefit 382–3, 384; income 386–7; sales 386, 387; see also property tax
Taxpayers’ Bills of Rights (TABORs) 292
Taylor, E. S. 185, 187
Taylor, L. L. 265
TCI (teacher cost index) 266–7
Teacher Advancement Program (TAP) 535
teacher cost index (TCI) 266–7
teacher cost indices from cost functions 267–8
teacher effectiveness: about 528–31; assessing 529–1; attracting effective novice teachers to low-performing schools 531–2; context 528–30; differential selection and retention of effective teachers 531–7; improving 537–9; incenting transfers for effective teachers 533; removing ineffective teachers 535–7; retaining effective teachers 534–5; teacher attrition in low-performing schools 533–4
teacher evaluation: about 174–5; classroom observations, standards-based 178–80; costs 185; delivery of 186; frameworks and components 178–84; future research 199; history 175–8; implementation considerations 184–5; measures, combining 184; outcomes 186–8; parent surveys 183–4; portfolios 184; purposes 174–5; reform efforts 177–8; removing ineffective teachers 536; student learning objectives 167–2; student surveys 183; teacher effectiveness and 529; teacher quality and 165, 167, 168n19; value-added measures 180–8
teacher quality: academic proficiency and 161–2; degree level and subject-matter preparation and 160–1, 168n13; deselection policies and 166–7; hedonic teacher cost index 266–7; hiring processes and 163–5; licensure exam test performance and 167; licensure status and 159–60; merit pay and 166, 168n21; NBPTS certification and 161, 165; professional development and 166; rural schools 611–12; school inputs in education production function 130; teacher characteristics and 158–62, 168n13; teacher evaluation and 165, 167, 168n19; teacher gateway policies and
teachers: bilingual 594, 602–3; charter school 428–9; cheating by 201; degree level and subject-matter preparation 160–1, 168n13; desegregation of schools 162–5; dual language programs 206–7; highly qualified teachers 167n4, 611–12; hiring processes 163–5; licensure exams, test performance on 162; licensure status 159–60; linguistic minority students and 594–5; low-performing schools, attracting to 531–3; low-performing schools, attrition in 533–4; pre-service preparation 537–8; professional development 166, 538, 612; recruitment and retention 534–5, 611, 612–14; removing ineffective 535–7; retaining 534–5; salaries 14–15, 166, 168n21, 529; strategically assigning 539; tax and expenditure limits and 480–1; tenure 536; transfer of 533, 536–7; working conditions 528, 534–5

teachers unions 33, 78, 172–9, 480, 491
Teach for America (TFA) 139, 167, 146, 532, 613–14
teaching, alternative pathways to 162–5, 532, 613–14
teaching experience, years of 160
teaching to the test 197–8, 201
technical efficiency 341
technology 65–6

television, interactive 610
TEIs see tax and expenditure limits (TEIs)
Tennessee 89, 129–30
Tenth Amendment 69
tenure, teacher 536
test cultures 115–16
tests, standardized 197–200, 201, 203, 544–7
Texas: accountability, school 202; charter schools 431; educational finance 46; equity and adequacy in school finance 234–5; intergovernmental aid formula 285; school finance litigation 54
TFA (Teach for America) 159, 162, 146, 532, 613–14
Theil index 278n4, 558, 558–9
Thomas, D. 521–7
Thomas B. Fordham Institute 83
Tiebout model of local public finance 218–19, 318
Timpane, P. M. 348
TIMSS (Trends in International Mathematics and Science Study) 322, 324–6, 332n7–8, 552
Title I of the Elementary and Secondary Education Act (1965) 322, 324–6
Title VII of the Elementary and Secondary Education Act (1965) 223–4
Toledo, Ohio 451
Tomaskovic-Devey, D. 608
Tomkiewicz, S. 513–14

total district revenue per pupil (TRPP) 284
transaction cost economics 32, 14–5

transfers, teacher 533, 536–7

transient student populations: about 623–5; absenteeism 626; policy interventions 634–5; resource and organizational implications 631–4; school accountability 208; schools and settings most likely to experience 630–1; social and academic impacts 623–4; students most likely to be mobile 628–30
transitional bilingual programs 291–2, 202
transition to college see high school-college divide
translog cost function 273n7
treatments 82
Trends in International Mathematics and Science Study (TIMSS) 306, 308, 109, 111, 112
TRPP (total district revenue per pupil) 284
TTI (Talent Transfer Initiative) 533
two-way immersion programs 592, 597
Tyler, J. H. 185, 187

591
Underwood, Julie 237
unions, teachers 74, 78, 177–8, 400, 401
University of Chicago 80–1
urban areas, transient student populations in 630
Urquiola, M. 454, 455
U.S. Census Bureau 338
U.S. Constitution 69, 75
n2
U.S. Department of Education: elementary and secondary expenditures 321, 326; Moving to Opportunity program 547, 548, 551; political context of education finance 74
U.S. Supreme Court: desegregation 323, 545; Equal Protection Clause 42, 45–6, 234–5; linguistic minority students 590; see also specific cases
user fees 305–11
validity 89
value-added models 132–3, 180–2
values, social 137–8
variables, instrumental 96–7, 99
venture philanthropy 345–6
Vergara litigation 54
Vermont 289, 467
Verstegen, D. A. 272–3, 615
vertical equity: as adequacy 239, 239–40; adequacy and 236–8; defined 232, 245; educational finance reform litigation 43; measures 246–7; as outcome equity 235–6; pure 239, 239 in tax policy 233; see also equity
vertically distributed power 69–70
Vigdor, J. L. 500, 550
virtual charters 427–8, 482–3
vocational education 320–1
voluntary contributors, locally-based 339–41
vouchers: about 16–17, 444–5; accountability 479; competitive effects 215, 226, 217–18, 479; current landscape 477–9; debate over 444; described 471–2; disadvantaged students 557, 558; domestic programs 447–53, 466–7, 469–7; for educational resource allocation 30, 31; effectiveness 31–2; future research 458–9; privately funded programs 450–1, 457; publicly funded programs 447–8, 468–7; state policies regarding 477–9; structure of 466–7; student achievement 458, 479; student demand for 478–9; student sorting and school competition 219, 220; see also charter schools
W
WADM (weighted average daily membership) 633
wage gaps 501–5
Walters, C. R. 435–6
Walton Family Foundation 344, 459
War on Poverty 324–6
Washington, DC see District of Columbia
Washington, George 78
"Washington Adequacy Funding Study" (Education Policy Improvement Center) 255
Washington state 255, 291, 447
waste 131–2
weak incentives problems 35
wealth 219
weighted average daily membership (WADM) 633
weighting formulas 287, 575–6
Weinberg, R. A. 513
West, M. R. 217, 554–5, 560n1
white flight 322–3
white students 430, 521–2; see also achievement gaps
whole-school design approach 250
Williamson, Oliver 34
Will It Be on the Test? (Johnson) 82–3
Willms, J. D. 221
Wilson, Duncan 303
Winters, M. A. 217, 554–5, 560n1
Wisconsin 467, 477
Wisconsin v. Yoder (1972) 480
Wise, Arthur 10, 43
Wolf, P. J. 32
workforce policies, teacher 165–7
working conditions for teachers 529, 534–5
World Bank 114, 115, 300; see also Education for All
World Declaration on Education for All 298, 299, 301
World War II 321–2
Wrobel, S. L. 414–17
Wyckoff, J. 188, 536, 539
Wyoming 50, 57n27, 290, 576

Y

Yinger, J. 897, 907–8
Youngs, P. 611

Z

Zimmer, R. 220