NCES
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement
NCES 2002–316

Developments in School Finance, 1999–2000

Fiscal Proceedings from the Annual State Data Conference July 1999 and July 2000

National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement
NCES 2002–316

Developments in School Finance, 1999–2000

Fiscal Proceedings from the Annual State Data Conference July 1999 and July 2000

July 2002

William J. Fowler, Jr.
Editor
National Center for Education Statistics

U.S. Department of Education
Rod Paige
Secretary

Office of Educational Research and Improvement
Grover J. Whitehurst
Assistant Secretary

National Center for Education Statistics
Gary W. Phillips
Deputy Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
1990 K Street NW
Washington, DC 20006–5651

July 2002

The NCES World Wide Web Home Page is: ***http://nces.ed.gov***
The NCES education finance World Wide Web Home Page is: ***http://nces.ed.gov/edfin***
The NCES World Wide Web Electronic Catalog is: ***http://nces.ed.gov/pubsearch***

Suggested Citation

U.S. Department of Education, National Center for Education Statistics. *Developments in School Finance, 1999–2000,* NCES 2002–316, William J. Fowler, Jr., Editor. Washington, DC: 2002.

For ordering information on this report, write:

U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

or call toll free 1–877–4ED–Pubs

Content Contact:

William J. Fowler, Jr.
202–502–7338
William.Fowler@ed.gov

Foreword

Jeffrey A. Owings, Associate Commissioner

Elementary/Secondary and Libraries Studies Division

At the 1999 National Center for Education Statistics (NCES) Summer Data Conference, scholars in the field of education finance addressed the theme, "Statistics, Technology, and Analysis for Tomorrow's Data Collections." Discussions and presentations focused on technology, data collection, and their implications on education finance reform. The theme for the 2000 Summer Data Conference was "Changing Data into Information: A Bridge to Better Policy" and focused on understanding data and survey changes, and again on their implications on education finance reform.

Developments in School Finance, 1999–2000 contains papers presented at the 1999 and 2000 annual NCES Summer Data Conferences. These Conferences attracted several state department of education policymakers, fiscal analysts, and fiscal data providers from each state, who are offered fiscal training sessions and updates on developments in the field of education finance. The presenters are experts in their respective fields, each of whom has a unique perspective or interesting quantitative or qualitative research regarding emerging issues in education finance. It is my understanding that the reaction of those who attended the Conference was overwhelmingly positive. We hope that will be your reaction as well.

This proceeding is the sixth education finance publication from NCES Summer Data Conferences. The papers included within present the views of the authors, and are intended to promote the exchange of ideas among researchers and policymakers. No official support by the U.S. Department of Education or NCES is intended or should be inferred. Nevertheless, NCES would be pleased if the papers provoke discussions, replications, replies, and refutations in future Summer Data Conferences.

Acknowledgments

The editor gratefully acknowledges the comments and suggestions of the reviewers, Tai Phan, Jeffrey Owings, and Marilyn McMillen Seastrom of the National Center for Education Statistics (NCES), and Mike Planty and Leslie Scott of the Educational Statistical Services Institute (ESSI). The following individuals from Pinkerton Computer Consultants, Inc. also contributed to the report: Susan Baldridge coordinated with each author to receive all materials and to resolve editorial issues, assisted in writing the introduction to the volume, designed the cover, and edited the manuscript; Carol Rohr designed the layout, performed the desktop publishing, assisted in editing the manuscript, and prepared the files for printing.

Contents

About the Editor

William J. Fowler, Jr.

National Center for Education Statistics

William J. Fowler, Jr. is the program officer of the Education Finance Program in the Elementary and Secondary Education and Library Studies Division at the National Center for Education Statistics (NCES), U.S. Department of Education. He specializes in elementary and secondary education finance and education productivity research. He is currently analyzing a fast-response survey for school business officials to obtain an understanding of how energy costs are affecting school district fiscal health. He is also engaged in research regarding kindergarten staff compensation. His great passion is designing Internet tools for the NCES education finance Web Site at http://nces.ed.gov/edfin, as well as the graphic display of quantitative data.

Dr. Fowler has worked for NCES since 1987, before which he served as a supervisor of school finance research for the New Jersey Department of Education. He has taught school finance at Bucknell University and the University of Illinois, and served as a senior research associate for the Central Education Midwestern Regional Educational Laboratory (CEMREL) in Chicago and for the New York Department of Education. He received his doctorate in education from Columbia University in 1977.

Dr. Fowler received the Outstanding Service Award of the American Education Finance Association (AEFA) in 1997, having served on its Board of Directors during the 1992–95 term, and has been re-elected for the 2001–04 term. He serves on the editorial board of the *Journal of Education Finance, Journal of Educational Considerations,* and the NCES *Education Statistics Quarterly*. He formerly served on the Board of Leaders of the Council for Excellence in Government, and was a 1997–98 Senior Fellow. He was a member of the Governmental Accounting Standards Board (GASB) Advisory Committee charged with developing a *User Guide for Public School District Financial Statements*.

Introduction and Overview

William J. Fowler, Jr.

National Center for Education Statistics

Education finance experts convened again in July of 1999 and July of 2000 for the annual National Center for Education Statistics (NCES) Summer Data Conference, a portion of which is devoted to presentations about public school finance. In each year, the focus of their discussions was theoretical perspectives on public school finance, the everyday policy concerns of schooling, and ongoing research studies. Each year, presenters are invited to submit papers based on the presentations made at the conference. This volume includes six papers from the 1999 and 2000 Summer Data Conferences from authors who accepted that invitation and addresses a variety of topics. They are intended to promote the exchange of ideas among researchers and policymakers. The views are those of the authors, and no official support by the U.S. Department of Education is intended or should be inferred.

The first paper analyzes the recent emphasis on performance-based accountability and asks the question: can we accurately define and measure school performance? One approach to developing school performance measures is to apply econometric and linear techniques that have been developed to measure productive efficiency. The study used simulated data to assess the adequacy of several of these methods for the purpose of performance-based school accountability. The results suggest that with the complex data sets and current technologies typical of education contexts, the most frequently used methods do not provide consistent measures of efficiency. Certainly this is an issue that Congress has recently struggled with in attempting to devise accountability measures for schools.*

The next three papers directly address education finance issues. One explores the use of national data to assess local school district spending on professional development. The authors rely on universe data from the U.S. Bureau of the Census' Survey of Local Government Finances: School District Finances (F-33). They discuss their techniques for blending the F-33 with the Common Core of Data, their use of Chambers' geographic cost adjustments (1998 version), their efforts to control for missing data, as well as choosing reporting statistics and interpreting those results into policy implications. Another paper explores the congressional mandate to the National Research Council's Committee on Education Finance to examine how education finance systems can be designed to ensure that all students achieve high levels of learning and that

* This assessment, based on information contained in *H.R. 1 Conference Report Highlights: Accountability for Student Achievement*, is available at http://edworkforce.house.gov/issues/107th/education/nclb/accountfact.htm.

education funds are raised and used in the most efficient and effective manner possible. The Committee's final report focused on the new challenges facing school finance. This volume includes the executive summary from the Committee's final report. The last paper reviews the policy shifts in education in the 1990s as standards-based reforms and accountability take hold. Specific analyses include the extent to which, and the ways in which, schools and school districts have adapted to reform pressure and opportunities.

The final two papers discuss evidence from litigation cases in various states and their effect on education finance. One paper asserts that court decisions in the Serrano tradition have undermined local support for education by creating property tax revolts, directly affecting the quality of their local schools. The other paper discusses lottery systems as revenue sources for public school finance. The author compares the results and reforms created by Georgia's lottery system to two neighboring states (Alabama and Tennessee) and their court-ordered education reforms. Specifically, the study addresses how these events influenced the size and composition of educational spending, as well as the distribution of the financial burdens.

Overall, the presentations at these conferences reiterated the age-old dilemma: what is the most effective and efficient revenue source and finance system and how are they best implemented? Let us now turn to the specifics of each paper.

In the first paper, *Evaluating School Performance: Are We Ready for Prime Time?,* **Robert Bifulco** and **William Duncombe** from the Maxwell School of Syracuse University assert that:

> There is a growing consensus that measures of school performance should be based on the student performance needs in the school. However, there is also recognition that any measure of school performance that is based on the performance of students needs to account for the differences in resources available to and service delivery environments faced by different schools.

There are a variety of econometric techniques to simultaneously consider school student performance, student needs, resources and staffing service configurations. Regardless of which quantitative technique is employed to estimate school efficiency, there are the twin questions of how accurate these estimates are, and how reliable they are. One would hope, for example, that regardless of economic method, the same schools would be identified as efficient schools. Bifulco and Duncombe use simulation to assess how accurately and reliably different econometric techniques identify efficient schools.

They are the first to acknowledge how "notoriously difficult" analysis of the education production function is. They assert that the first complication is that schools are charged not only with enhancing the cognitive development of students but also their social and emotional development, in order, for example, to promote democratic values. A second problem is the difficulty of measuring educational outputs. The use of standardized tests for cognitive attainment in particular academic subjects is problematic because such tests are not always aligned with curricular goals, and often do not assess higher-order thinking and problem solving. Then, of course, there is the major hurdle that very little is known about the factors that influence educational outputs. Even if some factors have been identified, such as teacher quality, measurement of such attributes is extremely difficult. Often, these techniques are also vulnerable to "unobserved variables," causing the mismeasurement of a school's efficiency. For example, if the activities of a teacher's union are responsible for higher teacher salaries, rather than higher teacher quality, and a measure of teacher unionism was not included in the econometric model, the resulting school efficiency score will be biased. Then, of course, there is also the dilemma that variables used by the models have simultaneous relationships. For example, poorly achieving schools may be in urban districts where their teachers are relatively inexperienced and less well educated. Finally, such environmental factors as student family background may overwhelm the school's effect on cognitive attainment.

Having acknowledged the difficulties, Bifulco and Duncombe examine six different econometric techniques for estimating school efficiency. They then conduct analyses in an attempt to determine whether these techniques are adequate for assessing school effectiveness within the context of school reform initiatives. They examine, for example, how well these econometric methodologies place schools in the lowest efficiency quintile. They found:

> In cases with endogeneity, measurement error, and a more complex production function, the best result was to place 31

> percent of the schools in the correct quintile. In these cases, at least 58 out of 200 were placed a quintile or 2 or more away from their true group. If such a method were relied on to determine financial awards or target corrective action, a large number of schools that lose out on additional resources or face burdensome requirements would have legitimate complaints. It also seems unlikely that analyzing the practices of groups identified as high or low performing by these methods would be very informative.

Bifulco and Duncombe thus conclude that the most commonly applied versions of econometric models do not provide adequate school measures of efficiency. They then suggest ways to improve efficiency measurements.

In the second paper, one of the recurrent questions in school finance is spending for the professional development of staff. **Kieran Killeen** of University of Vermont, **David H. Monk** of The Pennsylvania State University, and **Margaret L. Plecki** of the University of Washington examine this question, using NCES finance data, in part to report on the difficulties they encountered as they sought to make sense of the available data. In their paper, *Using National Data to Assess Local School District Spending on Professional Development*, they focus their attention when reviewing the literature on the data used in assessments of professional development activities. They then discuss the NCES data and the methods they employed, and their findings, and, in summary, discuss the kinds of data that are needed at the classroom, school and district level.

Killeen, Monk, and Plecki examined the extant research on teacher professional development and discovered that the most common staff professional development is conducted by school districts, rather than a college or university course. State monies for these purposes were typically controlled by the school district, and used for teacher in-service days, conferences, and workshops. The cost per regular classroom teacher in 1994 ranged between $1,755 and $3,259, using the salaries of district and school administrators, substitute teachers, and materials and supplies. Often not included in these estimates are salary increases earned by attending such in-service. A recently adopted popular strategy is to release students early on a regular basis, and have teachers engaged in professional development. This less-costly option (no teacher substitutes are needed) may have the adverse impact of reducing student instructional time.

Killeen, Monk, and Plecki used 2 years of data from the NCES Common Core of Data (CCD) and the U.S. Bureau of the Census' "Annual Survey of Local Government Finances (F-33)": 1991–92 and 1994–95. The school district finance data contain a variable, instructional staff support, which is composed of improvement of instruction services and educational media services. They found, using Chamber's 1998 geographic cost adjustment, that on average, in 1994–95, school districts spent about 2.8 percent of total expenditures on instructional staff support, or about $200 per pupil. The range was from about 2 to 8 percent. Most were between 2 and 5 percent. In per-pupil terms, spending on instructional staff support grew by 25 percent between 1992 and 1995. They found that spending increased with urbanicity.

Killeen, Monk, and Plecki conclude that the best opportunity to build a new data set exists in refinements to existing national surveys, such as the NCES Schools and Staffing Survey (SASS).

The third paper is a reprint of the executive summary of *Making Money Matter: Financing America's Schools* from the National Research Council's Committee on Education Finance, edited by **Helen F. Ladd** and **Janet S. Hansen**.

> A new emphasis on raising achievement for all students poses an important but daunting challenge for Policymakers: how to harness the education finance system to this objective….This report argues that money can and must be made to matter more than in the past if the nation is to reach its ambitious goal of improving achievement for all students.

In order to achieve this, the Committee asserts that finance decisions should be explicitly aligned with broad educational goals. Heretofore, finance policy focused primarily on the availability of revenues or disparities in spending, rather than funds needed to improve the educational system's performance. The emerging concept of funding adequacy is helpful in that it shifts the focus of finance policy from money received to how the funds are spent, with what outcomes. However, applying the adequacy concept at this stage in its infancy is an art, rather than a science, and misuse is possible. The Committee

warns that political pressures may result in specifying adequacy at so low a level as to trivialize the concept, or so high that it encourages higher spending. They also maintain that making money matter more requires supplementary finance strategies, such as aligning financial incentives and performance; investing the capacity of the educational system; and empowering schools and parents.

Little is understood regarding how funds can assist schools serving concentrations of disadvantaged students to raise student outcomes. The key question that was posed to the Committee was:

> How can education finance systems be designed to ensure that all students achieve high levels of learning and that education funds are raised and used in the most efficient and effective manner possible?

The Committee transformed this question into three goals for education finance systems:

1. Education finance systems should facilitate a substantially higher level of achievement for all students, while using resources in a cost-efficient manner.
2. Education finance systems should facilitate efforts to break the nexus between student background characteristics and student achievement.
3. Education finance systems should generate revenue in a fair and efficient manner.

The Committee recognizes that the system of U.S. education is highly decentralized and diverse, with the average public school district supported almost evenly between the state and local government. Despite school finance reforms initiated about 1970, U.S. education still remains dominated by large disparities in educational spending, although there is evidence that intra-state disparities have declined, inter-state disparities may have increased. With this background, the Committee evaluates a variety of policy options employing these three strategies, and weighs the evidence on how effective they are likely to be. Finally, the report draws attention to the nation's need for better and more focused education research to help strengthen schools and bring about substantial improvements in student learning.

In the fourth paper, the question of how school district resource allocations have changed over time in response to the standards-based reforms and the accountability movement of the last decade is addressed by **Jane Hannaway** and **Shannon McKay** of the Urban Institute, with **Yasser Nakib** of George Washington University. They use the NCES Common Core of Data (CCD) school district level information in concert with the Annual Survey of Local Government Finances (F-33) to discover national trends in resource allocation patterns, and to explore whether those states that engaged in extensive reform demonstrated different resource allocations from the national pattern.

Their research was designed to evaluate whether a shift in the requirements and demands on schools in the 1990s resulted in different patterns of resource allocations. They were particularly interested in whether school districts under high performance pressure shifted resources in response to that performance pressure. As Hannaway, McKay, and Nakib discover, while finance studies over the last 30 years have concentrated on equity issues and the distribution of funds to school districts, studies of the use of resources within school districts, and especially schools, have been rare. While some longitudinal analyses have been conducted, the expenditure categories have been too highly aggregated. Hannaway, McKay, and Nakib studied regular public school districts with enrollments greater than 200 students from 1992–97. They find that, adjusted for inflation, districts increased their total current expenditures from 1992 to 1997 by 7 percent. Proportionately more (than the national average) was spent on instruction, instructional support services, and school administration. Increases on pupil personnel services, driven by special education, demonstrated the largest proportionate increases. They also find that while district administration declined during this period, school administration increased. Hannaway, McKay, and Nakib find these results surprising since, despite reform pressure, districts were only making marginal increases in instructional area spending. Special education mandates, they speculate, drove pupil support service spending, rather than standards and accountability reform.

Hannaway, McKay, and Nakib then examine four reform states: Kentucky, Maryland, North Carolina, and Texas. Each of these states had an accountability system that rewarded high achieving schools, while differing in their financial status, with Kentucky increasing its expenditure levels by over four times the national average, and

Texas by more than twice the national average, particularly in instruction. Using multivariate analysis, Hannaway, McKay, and Nakib estimate the effect of being in a high reform state on school district resource allocation. Even after controlling for region, poverty level, urbanicity, and special education populations, and the level of spending in 1992, the researchers found Kentucky and Texas still increased investment in instruction more than Maryland and North Carolina. Hannaway, McKay, and Nakib interpret this result to mean that reform alone is insufficient for reallocation.

In the fifth paper, **William A. Fischel** of Dartmouth College gives his interpretation of school finance equity litigation and what he believes are subsequent property tax revolts. His argument is that court decisions that undermine local educational funding through the local property tax disconnect local funding and the educational quality of local schools. As a result, he believes the quality of public education in the United States has probably gotten worse as a result of school finance equity litigation. Fischel has written the paper in a nonacademic manner for policymakers, rather than fellow economists.

As noted by Fischel, there remains considerable variation in spending per pupil within most states, among states, and some poor urban school districts' conditions (as noted by Jonathan Kozol in *Savage Inequalities*) that are simply intolerable. The California Supreme Court was the first to insist on statewide funding equity (in 1971), and that at least 17 other state courts subsequently also have done so. By 1978, Fischel argues, taxpayers revolted with Proposition 13. In 1973, the U.S. Supreme Court decided the use of local property taxation to finance education did not violate the equal protection clause of the U.S. Constitution. Fischel also notes that there is inequality in school district property taxes, with wealthy school districts, such as Beverly Hills, raising more than twice as much revenue per student from its property than some poor school districts, even with a tax rate half as much as the poor school districts.

Fischel advances the argument that "unequal tax rates and tax bases are not themselves indicators of unequal economic burdens." To support this notion, he introduces us to the idea of "tax capitalization." A young economist named Charles Tiebout first proposed this idea in 1956. He believed that people could indicate their preference for a public service by "voting with their feet." Families, in short, will choose the best combination of housing and public services they desire. Zoning further enforces such choices. Although some large cities do not offer such mobility choices, Fischel argues that for most people, there are scores of different school systems from which to choose. As early as 1969, Wallace Oates confirmed that the prices of homes in communities with lower taxes or better services were higher. Fischel repeated such a study in New Hampshire in 1995, including tests scores given to fourth-graders. He concludes that school tax rates and test scores are "capitalized" in the value of owner occupied housing. The higher the properties tax, the lower the value of housing. Thus, Fischel argues, property tax rates do not measure the economic burden of the property tax system. He believes it is not unfair for houses in the low-tax town to have a higher price tag. "In the high-tax town, you pay more of your money to the tax collector; in the low-tax town, you pay more of your money to the mortgage banker." He also argues that the correlation between towns with high property wealth per pupil and towns with high median family income is low, and often negative. The reason, he believes, is that nonresidential commercial and industrial property often offsets low family income. "Accidents of geography" he asserts, are few and far between.

Finally, Fischel argues that local control over educational spending produces better educational results. Although he does not assert that the local property tax should be the only method of funding schools, he wishes to warn us that government intervention should be careful not to undermine the "virtues" of the local system. Homebuyers behave as if they know about the quality of local education. Competition among public schools, Fischel asserts, raises the quality of all. He then concludes that state court decisions requiring equality and higher state revenues have contributed to tax revolts. He then reviews the evidence in Maine, New Jersey, Massachusetts, Michigan, and New Hampshire.

In the last of the papers in this volume, **Thomas S. Dee** of Swarthmore College asks where new money goes, using evidence from successful state education litigation and a lottery in Georgia, Massachusetts, and Tennessee, (3 reform states) compared to Connecticut, Maine, and South Carolina. As discussed in the previous paper, state courts in 17 states have encouraged new aid to their poorest school districts. In addition, over the last 30 years, 37 states have also sought to enhance their education revenues with new state lotteries. What Dee asks is if either of these approaches results in more education funds, and

if so, where they go. In 1993, Massachusetts began court-order education finance reforms, increasing aid to poor school districts; Tennessee also did the same, and Georgia began a lottery to enhance education spending. Each state had its own unique strategy to assure that the funds enriched educational quality. Dee compares these three "reform" states with three neighboring "control" states: Connecticut, Maine, and South Carolina.

Dee explains that relative to school districts in the North, those in the South have less total revenue and receive less from local sources. He finds that real per-pupil state aid in Tennessee and Georgia increased following their 1993 reforms. However, he finds that it is difficult to untangle these increases from the recession recovery. Using regression, he finds that the reforms did increase state aid to schools in Massachusetts and Tennessee from $659 to $682 per pupil in state revenue.

In Georgia and Tennessee, school districts used their new aid to substantially reduce their outstanding debt. Approximately 53 percent of the new lottery-based revenue in Georgia went towards student instruction, while in Tennessee only 28 percent was so directed. Neither Georgia nor Tennessee's reforms, Dee finds, had the "intended consequence" of increasing school construction. However, the reforms did lead to increases in the purchases of instructional equipment.

A Final Note

Many readers are often unaware of the many conferences and training opportunities offered by the NCES in which the U.S. Department of Education pays most costs. Although there is the impression that state or local government employees may only attend these events, applications are for all those who utilize the NCES data.

The NCES routinely hosts several conferences annually. The annual Management Information Systems (MIS) conference is usually held in March of each year, and cosponsored with a state. The NCES Summer Data Conference, held in Washington, DC, usually the last week of July, is the source of the papers in this volume.

The NCES also offers training seminars that are open to advanced graduate students, researchers and policy analysts, and faculty members from colleges and universities. The 3- to 4-day seminars are usually held in the Washington, DC area, and are often specific to an NCES data set, such as the Early Childhood Longitudinal Survey (ECLS), or the Schools and Staffing Survey (SASS). Readers should check the NCES Web Site at http://nces.ed.gov/conferences/ for future conferences and training of interest.

Evaluating School Performance: Are We Ready For Prime Time?

Robert Bifulco

William Duncombe

Maxwell School of Citizenship and Public Affairs

Syracuse University

About the Authors

Robert Bifulco is a Ph.D. candidate in public administration at the Maxwell School of Citizenship and Public Affairs, Syracuse University, and a Research Associate at the Center for Policy Research. He has also worked for the New York State Education Department where he was involved in programs to identify and support low-performing schools.

William Duncombe is Associate Professor of Public Administration at the Maxwell School of Citizenship and Public Affairs, Syracuse University, and a Senior Research Associate at the Center for Policy Research. He has published numerous articles on school finance, the estimation of educational costs and the evaluation of educational programs.

Evaluating School Performance: Are We Ready For Prime Time?

Robert Bifulco

William Duncombe

Maxwell School of Citizenship and Public Affairs

Syracuse University

Introduction

Performance-based school reform has received much attention in recent years. Key elements of this reform movement include setting standards of student, teacher and school performance, granting autonomy to local actors in the educational process, and establishing rewards for high performance and remedies for low performance. These elements are prominently featured in the 1994 reauthorization of the Federal Title I program as well as several state-level reform initiatives.[1]

These reforms have been advanced as a remedy for several perceived problems with existing public education systems. Prominent among these perceived problems are a lack of the incentives and knowledge needed to improve student performance. Some have argued that given current systems for determining compensation, professional advancement and school funding, the incentives of school officials are insufficiently linked to student performance (Hanushek 1995, Levin 1997). Performance-based school reform attempts to provide stronger incentives for improving student performance by developing measures of achievement and tying financial and other rewards to those measures. Some also believe that we know very little about how to manage classrooms, schools and districts in ways that consistently result in higher levels of student achievement. By granting local actors the autonomy to experiment with new approaches and providing the means to assess the impact of local experiments on student performance, performance-based school reform is seen as a way to learn how to meet the ever-increasing demands placed on our public education systems (Hanushek 1995).

Developing valid and reliable measures of school performance is crucial both for efforts to establish incentives and to assess management practices. There is a growing consensus that measures of school performance should be based on the student performance needs in the school. However, there is also recognition that any measure of school performance that is based on the performance of students needs to account for the differences in resources available to and service delivery environments faced by different schools. One approach to developing such measures is to apply the concept of productive efficiency and techniques for measuring it, developed in the fields of economics and operations research.

[1] For examples and analysis of state level efforts, see Richards and Sheu (1992); Elmore, Abelmann, and Fuhrman (1996); and King and Mathers (1997).

Several such techniques have been developed, and several have been applied to estimate the efficiency of educational organizations. These include econometric approaches that utilize ordinary least squares regression and stochastic frontier estimation as well as a group of linear programming approaches falling under the rubric of Data Envelopment Analysis (DEA).[2] The availability of these methods for estimating school efficiency raises two questions. The first is whether or not the methods provide sufficiently accurate estimates of efficiency. The second question is, which method provides the most accurate estimates of efficiency and under what circumstances? Studies that have applied different methods to the same data have found that they provide different results (Banker, Conrad, and Strauss 1985; Nelson and Waldman 1986). The problem is that without knowing the true efficiency of the organizations studied, there is no way to determine which measures provide better estimates.

Studies that use simulated data with specified, and thus known, technological relationships and levels of efficiency can help to answer these questions. A limited number of such studies have been conducted, and recently some attempts have been made to use the results of such simulation studies to assess how appropriate existing efficiency measures are for the purposes of performance-based school reform. This paper reviews existing studies and provides new evidence from an analysis using simulated data.

…for any given combination of school inputs, a technically efficient school could not produce more of any output (holding the other outputs constant).

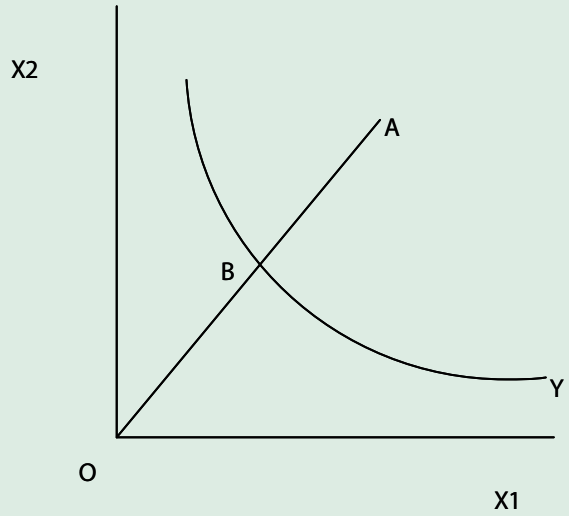
The body of this paper is presented in six sections. The first briefly describes the two general approaches to measuring productive efficiency used in the economic and operations research literature. The second section identifies the specific set of challenges that educational production processes pose for methods of estimating school efficiency. The third section reviews existing studies that have used simulated data to evaluate methods of estimating school performance. The fourth section identifies two different regression-based and four different DEA methods for measuring efficiency that we examine in a new simulation study. The fifth section describes how we simulated our data and the sixth section presents an analysis of how well each method did in estimating the known efficiencies of the simulated schools. The conclusion offers remarks concerning the current state-of-the-art in measuring school performance and the implications this has for performance-based school reform efforts.

Two Approaches to Measuring Productive Efficiency

Technical efficiency is defined as a feasible combination of inputs and outputs such that it is impossible to increase any output (and/or reduce any input) without simultaneously reducing another output (and/or increasing any other input). In other words, for any given combination of school inputs, a technically efficient school could not produce more of any output (holding the other outputs constant). The curve in figure 1 represents the combinations of inputs *X1* and *X2*, which if used efficiently, will produce *Y* units of output. This curve constitutes the efficient production frontier. The combination of inputs used by school 'A' to produce *Y* units of output places it off the efficient frontier. School A could produce *Y* units of output with less of either or both inputs. A measure of school A's technical efficiency can be calculated by dividing the length of ray *OB* by the length of ray *OA*. This measure represents the proportional amount of each input used by school A required to produce the level of output that it is producing. The fact that this measure is less than one indicates that school A is inefficient.

[2] Bessent and Bessent (1980) and Bessent et al. (1982, 1983) have applied the basic formulation of DEA developed by Charnes, Cooper, and Rhodes (1978) to schools in Houston. Färe, Grosskopf, and Weber (1989) have applied a version of DEA that allows for variable returns to scale to school districts in Missouri. More recently, Ray (1991); McCarty and Yaisawarng (1993); Ruggiero, Duncombe, and Miner (1995); and Kirjavainen and Loikkanen (1998) have applied DEA-based approaches that attempt to control for the different environmental factors faced by educational organizations. With regard to regression-based approaches, Barrow (1991), Deller and Rudnicki (1993), and Cooper and Cohn (1997) have applied stochastic frontier estimation methods to estimate the efficiency of districts, schools and classes. Stiefel, Schwartz, and Rubenstein (1999) have recently reviewed the various methods available for measuring school efficiency enumerating some of the advantages and disadvantages of each.

Figure 1.— Technical efficiency



Technical efficiency for firm A = OB / OA

SOURCE: Authors' sketch.

Econometric Approaches

The different methods for empirically estimating this measure of efficiency can be distinguished by the mathematical models used to estimate the efficient production frontier. Regression-based approaches begin by regressing an aggregate measure of output against a vector of inputs and a vector of environmental variables using ordinary least squares. Next, the estimated intercept term is "corrected" so that the estimated equation can be interpreted as a production frontier. The simplest of these methods, typically referred to as corrected ordinary least squares, increases the intercept term by the amount needed to make the largest residual zero. More complicated methods make use of assumptions about the probability distributions of inefficiency and random error to determine the intercept correction. Because the latter methods attempt to account for the affect of random, i.e., stochastic, factors on the observed relationships between inputs and outputs, they are said to estimate a stochastic production frontier.[3]

Implementing regression-based methods requires the assumption of an explicit functional form, explicit weights for each output and particular distributions of inefficiency and random error. The need to make assumptions that are difficult to verify is the primary disadvantage of these approaches to efficiency measurement. If the assumptions made are valid, however, the residual on the "corrected" regression equation for each school can be interpreted as a measure of inefficiency.[4]

[3] See Bauer (1990) for a review of a number of techniques for estimating stochastic frontiers.

[4] Stiefel, Schwartz, and Rubenstein (1999) discuss alternative regression-based measures of efficiency for cases where school-level panel data are available. Repeated observations on individual schools provided by panel data allows the estimation of school fixed-effects. Stiefel, Schwartz, and Rubenstein suggest that these fixed-effects may provide a better measure efficiency than the residual from cross-sectional ordinary least squares regressions because not all of the residual variation is attributed to efficiency. As the authors point out, however, a school fixed-effect reflects all systematic variation in outputs that is not explained by observed inputs, and therefore, is likely to reflect more than just differences in efficiency. In fact, because estimation of school fixed-effects precludes inclusion of time-invariant inputs in the regression equation, the fixed-effects reflects differences in these inputs as well as the impact of factors that are typically difficult to measure and include in cross-sectional regressions. Thus, it is doubtful that this use and interpretation of panel data estimates provides improved measures of efficiency, and may even be more misleading than measures based on the residual of cross-sectional estimators.

Data Envelopment Analysis

Linear programming techniques for estimating production frontiers fall under the rubric of Data Envelopment Analysis (DEA). All DEA methods start with measures of a set of inputs and outputs for some sample of schools. They then use numerical methods to select, for each school, the set of input and output weightings that maximizes the ratio of weighted outputs to weighted inputs. This maximization problem is subject to the constraint that the weights selected for a given school, when applied to other schools in the sample do not result in one of the other schools having a ratio of weighted outputs to weighted inputs greater than one. This maximum ratio of weighted outputs to weighted inputs is the measure of efficiency. The optimization problem is run for each school separately. Thus, each school will have a different set of input and output weightings. In effect, DEA selects the set of weights that will give a particular school as high an efficiency score as possible, subject to the constraint that no other school would have an efficiency score greater than one given those weights.[5]

DEA does not require *a priori* specification of output weights. Rather the linear programming procedure uses the data to determine relative output weightings for each school individually. Nor does it require assumptions about the functional form of the production frontier or the distribution of inefficiency. Given the uncertainty surrounding these aspects of educational production these are potentially important advantages. The primary disadvantage of DEA is that it is deterministic. That is, it attributes all deviation from the production frontier to inefficiency, and provides no means of accounting for random error.

Complications Posed by the Educational Production Process

Analysis of educational production is notoriously difficult.[6] Here, the focus is on aspects of education production that complicate the measurement of efficiency. The first difficulty is that education involves joint production of multiple outputs. Not only are schools charged with developing cognitive skills in several subject areas, but they are also charged with developing affective traits, promoting democratic values and furthering other social outcomes. Assumptions that these multiple outcomes are complimentary or even mutually consistent are difficult to maintain, and attempts to develop *a priori* weights that reflect the relative value of various outcomes are problematic. The fact that DEA does not require *a priori* specification of weights is typically touted as one of its primary advantages over regression-based approaches.

The second problem in analyzing educational production concerns the difficulty of measuring educational outputs. Standardized tests of cognitive skills are typically used to measure educational output. However, standardized tests are not always aligned with curricular goals, subjects such as science, social studies and the arts are not often tested, and even in tested subjects, higher-order thinking and problem solving skills are often not assessed (Darling-Hammond 1991). Valid and reliable measures of affective traits, democratic values and social outcomes may be even more difficult to obtain. The presence of this type of measurement error can push a school off the production frontier even if it is truly efficient or make it appear efficient when it is not. In so far as DEA attributes all deviation from the production frontier to inefficiency, its estimates of efficiency will be distorted by measurement error.

Analysis is further complicated by the fact that our knowledge about which factors affect educational outputs is inadequate.

Analysis is further complicated by the fact that our knowledge about which factors affect educational outputs is inadequate. In addition, measuring factors that are known to effect educational outputs, such as student motivation or teacher quality, can be difficult. Consequently, attempts to analyze educational production suffer from the presence of unobserved inputs. Because input levels are typically correlated with each other as well as with environmental factors, the problem of unobserved variables can cause the statistical estimation of model coefficients to

[5] The production frontier identified by DEA is a piecewise linear surface connecting each school that receives an efficiency rating of one. A school's efficiency score can be interpreted as its distance from this piecewise linear production frontier.

[6] For discussions of these difficulties, see Bridge, Judd, and Moock (1979) and Monk (1990).

be biased. This is will cause regression-based techniques to misplace the production frontier, thereby biasing measures of efficiency. More generally, a school's distance from the production frontier will be determined by variation on the omitted inputs, as well as inefficiency. Failure to account for this other source of variation in a school's distance from the estimated production function will lead both DEA and regression-based approaches to provide biased estimates of efficiency.

Education production is also characterized by simultaneous relationships between inputs and outputs. In the case of certain student inputs that affect the learning process this is clear. Student motivation, for instance, both influences and is influenced by the level of educational output. Orme and Smith (1996) suggest that there may be feedback from outputs to institutional inputs as well. School districts in which test scores are low might come under pressure to promote improved performance, which might lead to increased resource provision and thus higher levels of inputs. To some extent this process is institutionalized in legislative programs. The federal Title I program, for instance, targets significant amounts of funds to schools with large numbers of students who show low levels of achievement. Such feedback is also likely to bias the estimation of regression coefficients, and Orme and Smith argue that it can bias DEA estimates of efficiency as well.

Two recent studies… suggest that simple versions of regression and linear programming approaches are inadequate.

Finally, environmental factors, such as the family background of the students served by the school, can substantially influence the level of output that schools obtain. Environmental factors are conceptually different than production inputs because they are beyond the control of school officials. If environmental factors can be represented as simple additive terms in a school's production function, then it may be acceptable to treat them as another set of inputs. In this case, environmental factors might not significantly complicate the estimation of efficiency. If, however, these factors interact with controllable inputs and technologies in nonadditive ways, then incorporating environmental factors into efficiency analysis will be complicated.

Existing Simulation Studies

There have been several studies of both regression-based and DEA methods of estimating productive efficiency.[7] Most of these are concerned with frontier and efficiency estimation generally, and do not specifically ask whether or not a given method provides measures of efficiency that are accurate enough for the purposes of performance-based school reform. Are the estimates of efficiency provided by existing methods accurate enough to serve as a basis for awarding financial incentives or targeting remedial efforts? Can these methods help us determine what managerial and resource allocation practices help to foster improved performance? Two recent studies have examined these questions and suggest that simple versions of regression and linear programming approaches are inadequate.

Brooks (2000) examines a regression-based approach for developing adjusted performance measures for schools that is similar to simple regression-based measures of efficiency used in the productivity literature. He focuses on the effect that correlation between efficiency (or in his terms "merit") and school inputs has on the accuracy of adjusted performance measures. Examining the case of one observable input and one output, he finds that an increase in the correlation between efficiency and the observed input of 0.10 decreases the rank correlation between the adjusted performance measure and the schools true merit by 0.065. He also finds that as the random error associated with the production of student performance increases, the adjusted performance measure becomes even more inaccurate. In cases where the correlation between the input and "merit" is high (above 0.50) and random error is relatively large, he finds that the rank correlation between the adjusted performance measure and true merit will most likely be statistically indistin-

[7] Aigner, Lovell, and Schmidt (1977) and Olson, Schmidt, and Waldman (1980) have used simulated data to compare different econometric methods for estimating stochastic production frontiers. Orme and Smith (1996) have published simulation studies that examine particular properties of a single DEA method; Ruggiero (1996) and Ruggiero and Bretschneider (1998) have used simulated data to compare different linear programming models. For simulation studies comparing stochastic frontier estimators and DEA, see Gong and Sickles (1992) and Banker, Gadh, and Gorr (1993).

guishable from zero. In such cases, adjusted performance measures are unlikely to provide useful rankings of schools by performance.

Given what we know about the education production process, we expect correlation between inputs and efficiency of the kind examined by Brooks. Important unobserved factors, such as student and teacher motivation, are determined simultaneously with school efficiency. The more efficient a school, the higher student performance and consequently the more motivated teachers and students are to work harder, thereby improving efficiency. Also, higher levels of observed inputs, such as more teachers allowing reduced class sizes, may increase teacher and student motivation. If this is true, we can expect efficiency to be correlated with observed inputs. Thus, Brooks' findings suggest that simple regression procedures for developing efficiency measures are probably not adequate for performance-based school reform.

Bifulco and Bretschneider (2001) examine corrected ordinary least squares (COLS) and the Charnes, Cooper, and Rhodes (1978) formulation of DEA. They find that in cases with simultaneous relationships between inputs and outputs and with measurement error, the rank correlation between efficiency measures and true efficiency levels are no higher than 0.24. In these cases neither DEA nor COLS is able to place more than 31 percent of schools in the correct performance quintile. Schools assigned to the bottom 20 percent of performers are as likely to have actual performance levels above the median as in the bottom quintile. This confirms Brooks' suggestion that simple regression-based procedures provide inadequate measures of efficiency in educational contexts, and implies a similar conclusion for simple versions of DEA.

In the study presented below, we examine the accuracy of more sophisticated versions of regression-based and DEA methods for measuring efficiency. These methods are more complicated and thus more difficult to introduce into program and policy practice. However, they address some of the shortcomings of the simpler methods examined by Brooks (2000) and Bifulco and Bretschneider (2001). It is worth examining whether or not these improvements allow measures of school efficiency that are adequate for the purposes of performance-based accountability.

Methods Examined in This Study

We examine the performance of two regression-based methods of estimating efficiency and four versions of DEA. The first regression-based method we examine is COLS in which the intercept term from the ordinary least square regression is increased by the amount needed to make the largest residual zero. This is the method examined in Bifulco and Bretschneider (2001). The second regression-based method, referred to here as a stochastic frontier estimator (SFE), makes use of assumptions about the distributions of inefficiency and random error to determine the intercept correction. Details on this method are provided in Olson, Schmidt, and Waldman (1980).

The primary advantage typically advanced for regression-based approaches is their potential for addressing measurement error by treating efficient frontiers as stochastic phenomena.

The primary advantage typically advanced for regression-based approaches is their potential for addressing measurement error by treating efficient frontiers as stochastic phenomena. That is, these methods attempt to decompose the deviation of actual production from the estimated frontier into a component that is due to inefficiency and a component that is due to random error. The regression-based methods applied here, however, are not fully stochastic. The COLS method is entirely deterministic. In both adjusting the intercept and interpreting the residual from the adjusted regression equation, it is assumed that all deviation from the production frontier is due to inefficiency. The SFE method examined here uses assumptions about the distribution of inefficiency and random error in determining how much to adjust the intercept of the regression equations. Once the intercept is adjusted, however, deviations from the frontier are assumed to either be due entirely to inefficiency or entirely to random error. If an observation is on the efficient side of the frontier, all of the deviation from the frontier is assumed to be due to random measurement error. If an observation is on the inefficient side

of the frontier, all of the deviation is attributed to inefficiency.[8]

The first of four DEA methods examined, referred to here as DEA I, is the input minimizing formulation of the Charnes, Cooper, and Rhodes (1978) version of DEA. Following the practice typical of early applications of DEA (Bessent and Bessent 1980; Bessent et al. 1982, 1983), only those factors over which school officials have control are included as inputs. This method, which is the one examined in Bifulco and Bretschneider (2001), has been criticized for ignoring the impact of environment on outputs (Ruggiero 1996), a particularly important issue in education.

The other three DEA-based methods examined attempt to develop estimates of efficiency that control for the influence of environment on production outcomes. One of these methods, referred to in this paper as DEA II, attempts to control for environment by including it as an input in the standard DEA linear programming problem. Although this approach fails to recognize the important conceptual distinction between environmental factors and production inputs, it may provide a practical means of accounting for the environment in DEA estimates of efficiency.

In order to examine the performance of these methods, we generated 12 different data sets that incorporate various aspects of education production discussed previously.

Another approach is a two-stage method that uses regression in an attempt to separate those parts of the DEA estimates that are due to the effect of environment on output from those parts that are due to inefficiency. In the first stage of this method, DEA I is applied, using only discretionary inputs, to develop preliminary efficiency estimates. In the second stage, these preliminary efficiency estimates are regressed on environmental variables. An adjusted efficiency estimate is then computed for each observation by multiplying the coefficients from the second stage regression by the mean value of each environmental variable, and adding this to the observed regression residual. Methods similar to this have been applied to educational organizations by Ray (1991) and by Kirjavainen and Loikkanen (1998).[9] In this paper we referred to this method as "Two-Stage DEA."

The final method we examine is that developed by Ruggiero (1996). This method modifies the standard DEA programming problem to find the minimum level of inefficiency for a given school relative to other schools that face environments no better than the one it faces. Throughout this paper we refer to this as the Ruggiero approach.

Our Data Simulations

In order to examine the performance of these methods, we generated 12 different data sets that incorporate various aspects of education production discussed previously. Each data set consists of 200 simulated observations. The relationships between inputs and outputs (i.e., the technologies) underlying each of these data sets are described here.

A Cobb-Douglas Technology with Additive Environment

The first sets of simulated data were generated from the following system of underlying technological relationships:

$$y_1 = x_1^{0.25}\, x_2^{0.25}\, n^{0.50}\, e^{v1-u1} \qquad \textbf{(1a)}$$

$$y_2 = x_1^{0.20}\, x_2^{0.20}\, y_1^{0.20}\, n^{0.40}\, e^{v2-u2} \qquad \textbf{(1b)}$$

The first output (y_1) is related by a Cobb-Douglas production function to two inputs (x_1 and x_2), an environmental factor (n), a random component representing mea-

[8] This interpretation of the residual is admittedly *ad hoc*. A fully stochastic approach would decompose the residual that remains after the intercept correction into an efficiency and random component, again based on *a priori* assumptions about the probability distribution of these two error components. Jondrow et al. (1982) details the procedure for this decomposition. Arguments made by Ondrich and Ruggiero (1997) demonstrate that the measures of efficiency with and without the decomposition are ordinally equivalent.

[9] Both Ray (1991) and Kirjavainen and Loikkanen (1998) use a Tobit model to estimate the second stage regression. Given that the DEA estimate is truncated at the value of one this is appropriate. The application of "Two-Stage DEA" in this paper, uses OLS to estimate a general linear regression model for the second stage. In past work done by the authors, we have found that this simplification has little effect on the resulting efficiency estimates.

surement error (*v1*), and an inefficiency term (*u1*). The amount of the second output (y_2) is determined by the same two inputs, the same environmental factor, and the amount of the first output produced. The second equation also has a Cobb-Douglas form and includes its own measurement error (*v2*) and inefficiency terms (*u2*).

We randomly generated observations for each of the inputs from a uniform distribution on the interval (90, 110). We also randomly generated observations for the environmental variable from a uniform distribution on the interval (0, 200). Given the relative degree of standardization among schools with respect to input measures such as class size, and the large differences in environmental conditions, such as student poverty, faced by different schools, we might expect substantially more variation in the environmental variable than the input variables. The coefficients on both the input and environmental factors were chosen so that the effect of environment on output levels is large relative to the impact of the discretionary inputs. Again, this is intended as a rough approximation of real life educational production.

The coefficients on both the input and environmental factors were chosen so that the effect of environment on output levels is large relative to the impact of the discretionary inputs.

In each case, observations for the measurement error terms, *v1* and *v2*, were generated from a normal distribution, $N(0,\sigma_v^2)$. The value of σ_v was varied to generate three different data sets. In the first data-set, σ_v was set equal to zero to simulate a data-set with no measurement error. In the second and third cases, σ_v was set equal to 0.1 and 0.3 to simulate cases with 'small' measurement error and 'large' measurement error respectively.

The inefficiency terms, *u1* and *u2*, were each generated from a truncated normal distribution, $N(0,\sigma_u^2)$. The distribution was truncated by setting all negative values equal to zero. For both *u1* and *u2*, σ_u, was set equal to 0.3. The overall level of efficiency for each observation was calculated as follows. First, the observed values of y_1 and y_2 were computed from equations (1a) and (1b) and the randomly generated values of x_1, x_2, *n*, *v1*, *v2*, *u1*, and *u2*. Then, efficient values of *y1* and *y2* were generated for each observation by setting *u1* and *u2* equal to zero. An efficiency value was then calculated as follows:

$$Efficiency = [w_1(y1) + w_2(y2)] / [w_1(y1^*) + w_2(y2^*)] \quad (2)$$

Where *y1* and *y2* represent observed values, *y1** and *y2** represent efficient values, and w_1 and w_2 are weights that represent the relative importance of each output. In all cases, *w1* and *w2* are both set at 0.50. The mean efficiency values for the data sets with no, small and large measurement errors were 0.891, 0.892, and 0.897, respectively.

Cobb-Douglas Technology with Additive Environment and Endogeneity

The second set of samples was generated from the same technology as was just described with one exception. The observations for x_2 were replaced by observations linked to the inefficiency terms. Specifically, we used the following to generate observations of x_2:

$$x_2 = 95 + 12u1 + 12u2 + \varepsilon \quad (3)$$

Where ε is a normally distributed variable, $N(0, 9)$. This resulted in a distribution of x_2 similar to that generated in the above described data sets, and correlations between x_2 and the efficiency value ranging from –0.679 in the case with no measurement error to –0.671 in the case with large measurement error.

A negative correlation between inputs and efficiency values can be one of the by-products of the type of feedback from outputs to inputs discussed by Orme and Smith (1996). Correlation between the general composed error terms, *u1+v1* and *v2+u2*, can also be the result of omitted variables. Thus, incorporating this correlation into the simulated data allows us to explore the impact of such feedback processes or omitted variables, i.e., endogeneity, on efficiency measurement.

Cobb-Douglas Technology with Interactive Environment

The following productive relationships were used to simulate the third group of data sets:

$$y_1 = x_1^{(0.25+0.25n)} x_2^{(0.25+0.25n)} e^{v1-u1} \quad (4a)$$

$$y_2 = x_1^{(0.20+0.20n)}\, x_2^{(0.20+0.20n)}\, y_1^{0.20}\, e^{v2-u2} \quad \textbf{(4b)}$$

This system of equations is similar to (1a) and (1b) except that the environmental term enters into the equation nonlinearly. Here the environment affects the level of output by modifying the effect of each input. The same observations for x_1, x_2, *v1*, *v2*, *u1*, and *u2* that were generated for the first set of samples were used for these data sets. Observations for n were randomly generated from a uniform distribution on the interval (0,1). As in the above cases, samples with no, small and large measurement error were simulated from this underlying technology. The final three samples generated incorporate endogeneity into the above technology. This is done in the same way as described above.

Taken collectively, the 12 data sets simulated for this study incorporate several important aspects of educational production such as multiple outputs, environmental factors, measurement error, and endogeneity. This will allow us to examine how these aspects of educational production affect our ability to measure school efficiency.

An important component of performance-based school reform is identification of the highest and the lowest performing schools in a jurisdiction.

Results

Our discussion of the results is divided into four sections. First, we discuss the performance of the regression-based methods. Next, we discuss the performance of the DEA-based methods. Then, we compare the performance of the regression-based approaches with the linear programming methods. Finally, we present some analysis aimed at determining whether or not the most effective regression and DEA methods are adequate for the purposes of performance-based school reform initiatives.

The Performance of Regression-Based Methods

Regression-based methods require specification of assumptions concerning the functional form of technological relationships as well as the distribution of measurement error and inefficiency. In the multiple output case, assumptions also have to be made concerning the relative weighting of the various outputs. In applying COLS and the SFE to our simulated data we specify the same set of assumptions regardless of the data set being used. In some cases, these assumptions match the specifications of the true underlying technology, and in other cases they do not. Thus, we can see how misspecification affects the performance of regression-based procedures.

More specifically, in applying both COLS and SFE, we use a Cobb-Douglas functional form with the environment entering additively. In cases where this is in fact the functional form of the underlying technology, the regression models are well-specified. However, the data sets in which the environment enters interactively with the discretionary inputs, represent cases where the regression models are misspecified. In cases with endogeneity, a different type of misspecification is introduced.

The output weights used in forming the aggregate outcome measure used in the regressions were chosen to match those used to calculate the true efficiency value. In real situations these weights might in fact differ from school to school, and in any case, are difficult to specify. By matching the weights used in applying our estimators with those used in generating the true efficiency values, we ignore this potential difficulty in applying regression-based methods. This fact should be kept in mind when evaluating their performance.

Table 1 presents Kendall-Tau rank correlation coefficients between estimated efficiencies and true efficiency values. This measure captures the ability of each method to correctly rank observations. An important component of performance-based school reform is identification of the highest and the lowest performing schools in a jurisdiction. The highest performing schools can then be rewarded and corrective actions can be targeted to the lowest performing schools. Identifying groups of high and low performing schools can also be useful for determining whether certain management or resource allocation practices consistently lead to either higher or lower levels of performance. Thus, the ability of a method to correctly rank schools is an important criterion for assessing the usefulness of the methods for the purposes of performance-based school reform. A high rank correlation suggests that the measure performs well in identifying differential efficiency.

Table 1.—Rank correlations between estimated and true efficiency values*

	COLS	SFE	DEA I	DEA II	Two stage DEA	Ruggiero approach
Cobb-Douglas technology with additive environment						
Without endogeneity						
No measurement error	0.866	0.866	0.173	0.575	0.156	0.535
Small measurement error	0.835	0.835	0.154	0.192	0.144	0.128
Large measurement error	0.596	0.596	0.095	0.071	0.085	0.103
With endogeneity						
No measurement error	0.457	0.459	0.247	0.616	0.274	0.585
Small measurement error	0.484	0.487	0.239	0.362	0.264	0.215
Large measurement error	0.380	0.399	0.273	0.302	0.285	0.244
Cobb-Douglas technology with interactive environment						
Without endogeneity						
No measurement error	0.290	0.280	0.138	0.143	0.122	0.505
Small measurement error	0.264	0.262	0.141	0.138	0.133	0.118
Large measurement error	0.259	0.255	0.108	0.087	0.090	0.095
With endogeneity						
No measurement error	0.104	0.108	0.215	0.275	0.233	0.458
Small measurement error	0.099	0.104	0.213	0.264	0.237	0.203
Large measurement error	0.076	0.090	0.260	0.306	0.281	0.233

* Correlations are Kendall Tau-b statistics.

SOURCE: Authors' sketch.

In all cases, the rank correlations for COLS and SFE are virtually identical. Both methods use the same OLS estimates to determine the production frontier slope parameters, and differ from each other only in the way they adjust the intercept to locate the production frontier. Thus, this finding is expected, and confirms similar findings by Ondrich and Ruggiero (1997). This finding suggests that although COLS might provide different cardinal measures of efficiency than SFE (and in fact does),[10] it is equivalent to SFE as an ordinal measure.[11] Thus, in situations where only the ordinal ranking of schools are required, the simpler COLS method might be preferable.

The Performance of DEA-Based Methods

As expected, DEA I provides poor estimates of efficiency in all cases. This method does not control for the influence of the environment on production outcomes and thereby confounds the affects of inefficiency with the affects of environment. Somewhat surprisingly, regressing the estimates from DEA I against the environmental variable in the "Two-Stage DEA" does not substantially improve the performance of DEA I on the rank correlation criteria. This result may be due to a misspecification of the second stage regression model. We used OLS to esti-

[10] In results not reported here, the mean absolute difference between the true efficiency and the estimated efficiency scores were larger for SFE than for COLS in cases without measurement error, and considerably smaller for SFE than for COLS in cases with measurement error. This shows that although ordinally equivalent, the efficiency measures provided by these two methods are not cardinally equivalent.

[11] Ondrich and Ruggiero (1997) demonstrate that the ordinal measure of efficiency provided by COLS is equivalent to that provided by the fully stochastic frontier of Jondrow et al. (1982) as well.

mate a linear regression model. However, the distribution of DEA I efficiency estimates, which provide the dependent variable in the second stage regression, is truncated at 1. Thus, a Tobit model may be more appropriate.

In cases without measurement error, DEA II and the Ruggiero approach provide improved efficiency estimates. DEA II achieves improved estimates by including the environmental variable as an input in the standard DEA program. This improves matters only in cases where environment affects production in an additive fashion. The Ruggiero approach achieves improved estimates by modifying the DEA program so that each school is compared only to schools that face an environment no better than the one it faces. The Ruggiero approach achieves improved measures of efficiency when the environment affects production in an interactive way, as well as when environment enters additively. However, the performance of both DEA II and the Ruggiero approach is substantially undermined by the presence of measurement error.

Comparisons of Regression-Based Methods with DEA-Based Methods

The primary advantage typically touted for regression-based approaches is that they provide a means of handling measurement error. Regression-based approaches also provide well established means of controlling for the effect of environmental factors. However, we have seen that the ordinal measures of efficiency provided by COLS and SFE are equivalent. This raises doubts about the ability of SFE to separate measurement error from inefficiency in a truly informative way. Nonetheless, in cases where the SFE model is well-specified and measurement error is present, we might expect SFE to provide more accurate estimates of efficiency than DEA.

The primary advantage typically touted for regression-based approaches is that they provide a means of handling measurement error.

Regression-based methods of estimating efficiency require specification of a functional form for the production function. If this is misspecified, then the regression estimates, upon which the efficiency estimates are ultimately based, will be biased. DEA on the other hand constructs a piecewise linear production frontier. This is a highly flexible functional form that can approximate most actual technologies. Thus, we would expect DEA to provide better estimates of efficiency in cases in which the functional form of the COLS model is misspecified.

Correlation between inputs and inefficiency will also bias OLS estimates of production function coefficients. Orme and Smith (1996) argue that the presence of such correlation can also bias DEA efficiency estimates. However, Bifulco and Bretschneider (2001) do not find any support for Orme and Smith's argument. Thus, we might expect the presence of endogeneity to have a larger impact on the performance of COLS and SFE efficiency estimates than DEA estimates

These expectations are by and large confirmed by the results reported in table 1. In cases where the COLS model is well-specified (i.e., the first three rows of table 1), COLS and SFE performs better than DEA on the rank correlation criteria. DEA tends to perform better in cases where the regression model is misspecified. In cases with endogeneity but no measurement error, DEA II and the Ruggiero approach both outperform COLS and SFE. When the functional form of the regression model is also misspecified (i.e., the last three rows of table 1), the Ruggiero method achieves higher rank correlations than COLS and SFE.

In cases where measurement error is present and the regression model is misspecified, both DEA and the regression-based methods perform poorly. It might be argued that these are the conditions most likely to be encountered in attempts to measure the efficiency of educational organizations. In these cases, the presence of measurement error substantially diminishes the performance of DEA, and the combination of measurement error and misspecification significantly diminishes the performance of COLS. In cases with measurement error, endogeneity and misspecified functional forms, rank correlations higher than 0.306 are never achieved, and in half of these cases rank correlations are below 0.15. It is doubtful that rank correlations of this magnitude are adequate for the purposes of awarding performance bonuses or targeting remedial resources.

Adequacy of Efficiency Estimates for Purposes of Performance-Based Reform

The results reported in table 1 raise doubts about whether our ability to estimate school efficiency is adequate for the purposes of performance-based school reform. To investigate this issue further we divided the observations in each of the 12 data sets into quintiles based on their true efficiency score. We then examined the ability of the most effective methods in the above analyses—SFE, DEA II, and the Ruggiero approach—to place observations in the appropriate quintiles. We also examined the true efficiency rankings of the schools identified by these methods as being in the lowest efficiency quintile. The results of these analyses are presented in tables 2 and 3.

The SFE method did well in cases where the underlying regression model was well-specified (see first three rows of tables 2 and 3). Particularly in cases with low measurement error, SFE assigned 74 percent of schools to the appropriate quintile and only 2 out of 200 schools were assigned to a quintile 2 or more away from their true quintile. The method also did well identifying the lowest performing schools. Of the schools assigned to the bottom quintile by SFE, 95 percent were actually in the bottom efficiency quintile and none of the schools had true efficiency values that ranked them higher in efficiency than the median.

However, the SFE method did not do as well in cases where the underlying regression model is misspecified.

Table 2.—Measures of how well various methods do in assigning observations to quintiles

	SFE		DEAII		Ruggiero approach	
	Percent assigned to correct quintile	Percent assigned two or more quintiles from actual	Percent assigned to correct quintile	Percent assigned two or more quintiles from actual	Percent assigned to correct quintile	Percent assigned two or more quintiles from actual
Cobb-Douglas technology with additive environment						
Without endogeneity						
No measurement error	80.0%	0.0%	36.0%	19.0%	34.0%	26.5%
Small measurement error	74.0	1.0	27.5	42.0	22.0	37.0
Large measurement error	49.0	11.5	20.5	43.5	21.0	38.0
With endogeneity						
No measurement error	38.5	24.0	41.5	15.5	39.0	24.5
Small measurement error	41.5	19.5	30.5	30.0	26.5	35.5
Large measurement error	34.0	28.0	32.0	29.0	26.5	36.5
Cobb-Douglas technology with interactive environment						
Without endogeneity						
No measurement error	26.5	38.5	26.5	41.0	35.0	29.5
Small measurement error	26.5	37.5	24.5	41.5	22.5	37.5
Large measurement error	28.5	38.0	20.5	45.0	22.5	37.5
With endogeneity						
No measurement error	27.5	46.0	30.5	30.0	33.5	32.0
Small measurement error	27.0	47.5	31.0	30.5	26.5	35.5
Large measurement error	24.5	46.0	31.0	29.0	25.0	36.5

SOURCE: Authors' sketch.

Table 3.—Measures of how well various measures do in identifying low efficiency schools

	SFE		DEAII		Ruggiero approach	
	Percent assigned to bottom quintile actually in bottom quintile	Percent assigned to bottom quintile actually ranked above median	Percent assigned to bottom quintile actually in bottom quintile	Percent assigned to bottom quintile actually ranked above median	Percent assigned to bottom quintile actually in bottom quintile	Percent assigned to bottom quintile actually ranked above median
Cobb-Douglas technology with additive environment						
Without endogeneity						
No measurement error	97.5%	0.0%	72.5%	0.0%	67.5%	0.0%
Small measurement error	95.0	0.0	45.0	35.0	30.0	47.5
Large measurement error	80.0	0.0	37.5	42.5	27.5	45.0
With endogeneity						
No measurement error	67.5	5.0	72.5	0.0	72.5	0.0
Small measurement error	72.5	5.0	47.5	12.5	35.0	32.5
Large measurement error	62.5	12.5	50.0	22.5	37.5	30.0
Cobb-Douglas technology with interactive environment						
Without endogeneity						
No measurement error	57.5	10.0	45.0	37.5	77.5	0.0
Small measurement error	57.5	10.0	40.0	45.0	30.0	47.5
Large measurement error	57.5	10.0	35.0	42.5	27.5	45.0
With endogeneity						
No measurement error	40.0	37.5	42.5	22.5	62.5	0.0
Small measurement error	37.5	35.0	42.5	22.5	35.0	32.5
Large measurement error	35.0	37.5	52.5	22.5	32.5	32.5

SOURCE: Authors' sketch.

In cases where the functional form is misspecified, SFE places more schools in quintiles two or more away from their true quintile than it places in the correct quintile. In cases where endogeneity is present and the SFE model is misspecified, less than half of the schools identified as being among the schools with the lowest level of efficiency are actually in the bottom efficiency quintile, and at least 14 of the 40 schools placed in the bottom quintile have true efficiency values that rank them above the median.

The two DEA methods did not do well in placing students in the correct quintile in any of the data sets. In no case did either DEA II or the Ruggiero approach place as many as half the schools in the correct quintile. In the majority of cases these methods place as many or more schools in a quintile two or more away from the true quintile as they place in the correct quintile. The DEA II and the Ruggiero approaches did reasonably well identifying low-efficiency schools, but only in the unrealistic cases where there was no measurement error.

It appears that if the underlying regression model is well-specified, then the SFE method can provide efficiency estimates that are adequate for at least some purposes. However, the past 35 years of experience in trying to analyze educational production suggest that the functional relationships between educational outcomes, school inputs and environmental factors are complex and that we know little about the forms these relationships take. Un-

fortunately, SFE does much worse when the underlying regression model is misspecified. The DEA method has been advanced as a method of estimating efficiency that does not depend on restrictive assumptions about the form of productive relationships. However, the estimates of efficiency provided by DEA, particularly in the presence of measurement error, do not appear to be adequate.

Whether the performance of SFE when the underlying regression model is misspecified or the DEA methods in the presence of measurement error is adequate for the purposes of school-based reform is a matter of judgment. However, it is difficult to argue that the results in tables 2 and 3 are adequate. In cases with endogeneity, measurement error, and a more complex production function (i.e., the last three rows), the best result was to place 31 percent of the schools in the correct quintile. In these cases, at least 58 out of 200 schools were placed a quintile 2 or more away from their true group. If such a method were relied on to determine financial awards or target corrective action, a large number of schools that lose out on additional resources or face burdensome requirements would have legitimate complaints. It also seems unlikely that analyzing the practices of groups identified as high or low performing by these methods would be very informative. If less than half of the schools that are identified as low performing are actually inefficient, and 30 percent are actually achieving above average levels of efficiency, then it is difficult to say that the managerial practices or patterns of resource allocation found in those schools are ineffective.

The COLS and SFE methods perform well in cases where the underlying model is well-specified, particularly when measurement error is small.

Conclusions

Existing studies as well as the new evidence presented here suggest that for the complex production processes found in schools, i.e., processes characterized by complex functional forms, endogenous relationships between inputs and outputs, and substantial measurement error, the most commonly applied versions of DEA and regression-based methods do not provide adequate measures of efficiency. It would be difficult to defend implementing performance-based financing or management programs with estimates of school performance whose rank correlation with true performance is no higher than 0.30. However, our results need not be interpreted with unequivocal gloom. Not only must our findings be properly qualified, but they also suggest strategies for developing more adequate measures of efficiency.

The COLS and SFE methods perform well in cases where the underlying model is well-specified, particularly when measurement error is small. DEA also performs much better in cases without measurement error. Some forms of DEA, particularly Ruggiero's approach, also appear to be fairly robust with respect to the functional relationships between outcomes, inputs and the environment. This suggests at least three avenues for improving efficiency measurement.

First, efforts to reduce the amount of measurement error characteristic of current educational data sets are needed. Such efforts are well under way. The 1994 reauthorization of the Elementary and Secondary Education Act provided substantial amounts of funding to state educational agencies to develop testing programs that are aligned with explicit curricular goals, that test higher level thinking skills and that can be used for purposes of evaluating school performance. States, such as Kentucky, have led the way in the development of such assessment systems.[12] In addition, several city school districts, including Chicago and New York City, have developed school-based budgeting systems. These systems provide more reliable school-level resource data than has ever before been available (Rubenstein 1998; Iatarola and Stiefel 1998).

In addition to reducing measurement error, it might be possible to modify existing methods of estimating efficiency so as to minimize the effect of measurement error and/or endogeneity. For instance, the fact that the performance of COLS and SFE is diminished by correlation between inputs and inefficiency is not surprising.

[12] See *Education Week*'s *Quality Counts 2001* publication entitled *A Better Balance: Standards, Tests and the Tools to Succeed* for a discussion of state efforts to develop improved assessments of student performance.

This type of correlation violates the assumptions that are required if ordinary least squares is to provide unbiased coefficient estimates. Bias in these coefficient estimates is the source of the poor performance of COLS and SFE in estimating efficiency. There are, however, well known simultaneous equation methods, such as two-staged least squares, that provide unbiased coefficient estimates in cases where the assumptions of ordinary least squares are violated. If such methods could be used to estimate production frontiers, then efficiency estimates that perform better than those we have examined might be developed.

Finally, efforts to understand the functional forms that characterize educational production are needed. These efforts may be the most important for improving efficiency measurement and the most difficult to achieve. However, with continued efforts to develop theory and test those theories with more complex empirical models, we may be able to make progress on this front. The use of flexible functional forms, such as the translog production function, might also help provide more accurate estimates of efficiency by relaxing some of the restrictive assumptions about production technology made in typical regression models.[13]

In addition, we must not overlook the possibility of augmenting quantitative measures of efficiency with qualitative forms of evaluation. Such qualitative forms of evaluation might involve site visits and audits by professional peers. Research is needed to determine exactly how information acquired through such methods can be combined with existing data and methods to develop more reliable and valid measures of school performance.

Given the data that are currently available, however, our results suggest that the methods for measuring the efficiency of educational organizations that have been used most frequently may not be adequate for use in implementing performance-based management systems. This is a discouraging result, and suggests that efforts to implement performance-based school reforms should proceed with caution.

[13] See Beattie and Taylor (1985) for details on the use of flexible functional forms in production analysis.

References

Aigner, D., Lovell, C.A.K., and Schmidt, P. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics.* 6: 21–37.

Banker, R.D., Conrad, R.F., and Strauss, R.P. 1985. "A Comparative Application of DEA and Translog Methods: An Illustrative Study of Hospital Production." *Management Science.* 32: 30–44.

Banker, R.D., Gadh, V.M., and Gorr, W.L. 1993. "A Monte-Carlo Comparison of Two Production Frontier Estimation Methods: Corrected Ordinary Least Squares and Data Envelopment Analysis." *European Journal of Operational Research.* 67:332–343.

Barrow, M.M. 1991. "Measuring Local Education Authority Performance: A Frontier Approach." *Economics of Education Review.* 10:19–27.

Bauer, P.W. 1990. "Recent Developments in the Econometric Estimation of Frontiers." *Journal of Econometrics.* 46: 39–56.

Beattie, B.R. and Taylor, C.R. 1985. *The Economics of Production*. New York, NY: John Wiley & Sons.

Bessent, A. and Bessent, E. 1980. "Determining the Comparative Efficiency of Schools Through Data Envelopment Analysis." *Educational Administration Quarterly.* 16:57–75.

Bessent, A., Bessent E., Charnes, A., Cooper, W., and Thorogood, N.C. 1983. "Evaluation of Educational Program Proposals by Means of DEA." *Education Administration Quarterly.* 19: 82–107.

Bessent, A., Bessent W., Kennington, J., and Reagan, B. 1982. "An Application of Mathematical Programming to Assess Productivity in the Houston Independent School District." *Management Science.* 28: 1,355–1,367.

Bifulco, R. and Bretschneider, S. 2001. "Estimating School Efficiency: A Comparison of Methods Using Simulated Data." *Economics of Education Review.* 20: 417–429.

Bridge, G., Judd, C., and Moock, P. 1979. *The Determinants of Educational Outcomes: The Impact of Families, Peers, Teachers and Schools*. Cambridge, MA: Ballinger Publishing Company.

Brooks, A.C. 2000. "The Use and Misuse of Adjusted Performance Measures." *Journal of Policy Analysis and Management.* 19: 323–329.

Charnes, A., Cooper, W.W., and Rhodes, E. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research.* 2: 429–444.

Cooper, S.T. and Cohn, E. 1997. "Estimation of a Frontier Production Function for the South Carolina Educational Process." *Economics of Education Review.* 16: 313–327.

Darling-Hammond, Linda. 1991. "Accountability Mechanisms in Big City School Systems." *ERIC/CUE Digest.* 17.

Deller, S. and Rudnicki, E. 1993. "Production Efficiency in Elementary Education: The Case of Maine Public Schools." *Economics of Education Review.* 12: 45–57.

Elmore, R., Abelmann, C., and Fuhrman, S. 1996. "The New Accountability in State Education Reform: From Process to Performance." In H.F. Ladd (Ed.), *Holding Schools Accountable: Performance-Based Reform in Education.* Washington, DC: The Brookings Institution. 65–98.

Färe, R., Grosskopf, S., and Weber, W. 1989. "Measuring School District Performance." *Public Finance Quarterly.* 17: 409–428.

Gong, B. and Sickles, R.C. 1992. "Finite Sample Evidence on the Performance of Stochastic Frontiers and Data Envelopment Analysis Using Panel Data." *Journal of Econometrics.* 51: 259–284.

Hanushek, E. 1995. *Making Schools Work: Improving Performance and Controlling Costs.* Washington, DC: The Brookings Institution.

Iatarola, P. and Stiefel, L. 1998. "School-Based Budgeting in New York City: Perceptions of School Communities." *Journal of Education Finance.* 23: 557–576.

Jondrow, J., Lovell, C.A.K., Materov, I.S., and Schmidt, P. 1982. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model." *Journal of Econometrics.* 19: 233–238.

King, R. and Mathers, J. 1997. "Improving Schools Through Performance-Based Accountability and Financial Rewards." *Journal of Education Finance.* 23:147–176.

Kirjavainen, T. and Loikkanen, H.A. 1998. "Efficiency Differences of Finnish Senior Secondary Schools: An Application of DEA and Tobit Analysis." *Economics of Education Review.* 17: 377–394.

Levin, H.M. 1997. "Raising School Productivity: An X-Efficiency Approach." *Economics of Education Review.* 16: 303–311.

McCarty, T. and Yaisawarng, S. 1993. "Technical Efficiency in New Jersey School Districts. In H. Fried, K. Lovell, and S. Schmidt (Eds.), *The Measurement of Productive Efficiency.* London, England: Oxford University Press. 271–287.

Monk, D.H. 1990. *Educational Finance: An Economic Approach.* New York, NY: McGraw-Hill Publishing Company.

Nelson, R.A. and Waldman, D.M. 1986. *Measuring Technical Efficiency: Index Numbers vs. Parametric Production Frontiers*. Unpublished Manuscript.

Olson, J.A., Schmidt, P., and Waldman, D.A. 1980. "A Monte-Carlo Study of Estimators of Stochastic Frontier Production Functions." *Journal of Econometrics.* 13: 67–82.

Ondrich, J. and Ruggiero, J. 1997. "Efficiency Measurement in the Stochastic Frontier Model." *Mimeo.* Syracuse University.

Orme, C. and Smith, P. 1996. "The Potential for Endogeneity Bias in Data Envelopment Analysis." *Journal of the Operational Research Society.* 47: 73–83.

Ray, S.C. 1991. "Resource Use in Public Schools: A Study of Connecticut." *Management Science.* 37: 1,620–1,628.

Richards, C. and Sheu, T. 1992. "The South Carolina School Incentive Reward Program: A Policy Analysis." *Economics of Education Review.* 11: 71–86.

Rubenstein, R. 1998. "Resource Equity in the Chicago Public Schools: A School-Level Approach." *Journal of Education Finance.* 23: 468–489.

Ruggiero, J. 1996. "On the Measurement of Technical Efficiency in the Public Sector." *European Journal of Operational Research.* 90: 553–565.

Ruggiero, J. and Bretschneider, S. 1998. "The Weighted Russell Measure of Technical Efficiency." *European Journal of Operations Research.* 108: 438–451.

Ruggiero, J., Duncombe, W., and Miner, J. 1995. "On the Measurement and Causes of Technical Inefficiency in Local Public Services: With an Application to Public Education." *Journal of Public Administration Research and Theory.* 5: 403–428.

Stiefel, L., Schwartz, A., and Rubenstein, R. 1999. "Measuring School-Level Efficiency." In Goertz and Odden (Eds.), *School Based Budgeting.* Thousand Oaks, CA: Corwin Press, Inc. 40–75.

Using National Data to Assess Local School District Spending on Professional Development

Kieran M. Killeen
University of Vermont
David H. Monk
The Pennsylvania State University
Margaret L. Plecki
University of Washington

About the Authors

Kieran M. Killeen is an assistant professor of educational leadership and policy studies at the University of Vermont and recently received his Ph.D. from Cornell University. His areas of research include public school finance, planning, and institutional theory. He is a former special education teacher and participant of the Teach For America national teacher corps. Previously, he was a lecturer at The Pennsylvania State University.

David H. Monk is a professor of educational administration and the dean of the College of Education at The Pennsylvania State University. He earned his Ph.D. in 1979 at the University of Chicago and was a member of the Cornell University faculty for 20 years prior to becoming dean at Penn State. He has also taught in a visiting capacity at the University of Rochester and the University of Burgundy in Dijon, France.

Margaret L. Plecki is an associate professor in educational leadership and policy studies at the University of Washington, Seattle. Her areas of specialization are school finance, education policy, and educational leadership. During her career, she has also served as a special education teacher, a school administrator, and a management consultant in the private sector. Dr. Plecki received her Ph.D. from the University of California at Berkeley.

Using National Data to Assess Local School District Spending on Professional Development

Kieran M. Killeen
University of Vermont

David H. Monk
The Pennsylvania State University

Margaret L. Plecki
University of Washington

Introduction

The limited progress that has been made over the years toward understanding the nature of relationships between resources and the learning gains of students is a source of great and recurring frustration for educational researchers, policymakers, and practitioners alike. Part of the difficulty can be traced to fundamental inadequacies in the data that speak to these and related issues. One of the areas where the data are particularly lacking concerns the area of spending for professional development. The limits of these data have seriously curtailed analysts' efforts to measure and assess investments in teacher professional development.

The purpose of this report is to describe procedures we used to analyze the available national data in our studies of professional development expenditures at the local school district level. We found that the *Annual Survey of Local School District Finances* or F-33[1] provides a rich set of school district level revenue and expenditure data. With some modification, the F-33 can provide the national perspective on a host of detailed revenue and expenditure items. Additionally, the F-33 is easily linked with other national data sets like the Common Core of Data (CCD). A closely related purpose of this study is to report on the difficulties we encountered as we sought to make sense of the available data. We have reported more detailed versions of our empirical findings elsewhere.[2] Our purpose here is to explore the data collection issues in greater detail than was possible earlier, and to provide an overview of the basic findings.

Our report is divided into four major sections. We begin with a review of the research literature dealing with spending on professional development and focus our attention on the data used in assessments of professional development activities. We turn next to a discussion about the data and the methods we employed in our work with the F-33. Finally, we summarize our general findings from two separate analyses of the F-33. In our summary and conclusion, we discuss the kinds of data that are needed

[1] F-33 is the actual name of the survey instrument, coded by the U.S. Bureau of the Census.

[2] See Killeen, K., Monk, D., and Plecki, M. 2000. "Spending on Instructional Staff Support Among Big City School Districts: Why are Urban Districts Spending at Such High Levels?" *Educational Considerations*. 28(1). Killeen, K., Monk, D., and Plecki, M. 1999. "School District Spending on Professional Development: Insights from National Data." A Working Paper of the Center for the Study of Teaching and Policy, University of Washington.

at the classroom, school, and district level learning gains for pupils.

Current Research on Professional Development Financing

While the literature on professional development financing is not extensive, several research efforts have sought to estimate the levels of investment in teacher professional development. Most of the research on investments in professional development has addressed the source, type, and/or amount of professional development purchased (Moore and Hyde 1981; Lytle 1983; Stern, Gerritz, and Little 1989; Elmore 1997; Education Commission of the States 1997). One study found that teachers are two to three times more likely to be participants in district-provided staff development than enrolled in a college or university course (Little 1989). The same study also calculated that more than four-fifths of state dollars for staff development were controlled by the local district. Professional development activities have been dominated by a training-based delivery system, generally managed by school districts, which offers teachers a variety of workshops targeted on special projects or narrowly defined aspects of reform (Little 1993). A study by the Education Commission of the States (1997) found that approximately three-fourths of school district resources designated for professional development are spent on teacher in-service days, conferences, and workshops.

…teachers are two to three times more likely to be participants in district-provided staff development than enrolled in a college or university course.

Miller, Lord, and Dorney's (1994) estimates of expenditures on staff development range between 1.8 and 2.8 percent of the district's operating budget. The cost per regular classroom teacher ranged between $1,755 and $3,259. Their study was based on a series of intensive case studies in four districts located in different regions in the United States, ranging in size from 9,500 to 125,000 students. The estimates are based on direct costs such as the salaries of district and school administrators, and substitute teachers, as well as on the direct costs of materials and supplies. One detailed study of staff development in California (Little et al. 1987) estimated the investment in professional development to be almost 2 percent of total funding for education in that state. In a study of one New York school district, Elmore (1997) estimated that spending on professional development amounted to about 3 percent of the total budget. These studies do not consider, however, that most districts, due somewhat to the requirements of the bargained contracts with teachers, compensate teachers for staff development activities through an increase in salary, thus representing a "hidden" cost of traditionally delivered staff development. For example, a study of spending on professional development in the Los Angeles Unified School District (Ross 1994) found that the district expended $1,153 million in teacher salaries in 1991–92, and that 22 percent of this figure could be attributed to salary point credits that were earned because of courses or other approved professional development activities on the part of teachers.

As the example of investing in professional development through salary increments implies, there is a pronounced difficulty in fully accounting for all staff development costs. Professional development activities frequently are financed through a combination of revenue sources, including nongovernmental sources, thereby complicating the cost accounting. Professional development experiences also might be associated with substantial contributions of volunteer time on the part of teachers (Little et al. 1987). At the same time, teachers might accrue additional credits for professional development activities, which advance them on the salary schedule, resulting in a long-term fiscal obligation to the district in the form of the resultant base salary increase. Finally, similar professional development activities might vary significantly in costs per teacher depending on the financing strategy that is employed. For example, one strategy for supporting teacher professional development that is increasing in popularity is the "early release" option in which students are released from school on some regular basis, thereby allowing time during regular school hours for teachers to engage in professional development. This option clearly is less costly for school districts, as it removes the additional costs of substitutes or additional hours worked by teachers. However, there is a significant opportunity cost borne by students in the form of reduced instructional time.

The studies of professional development costs briefly reviewed above concentrate on the more traditional forms of professional development delivery. However, significant changes have been taking place in recent years regarding the conceptualization of effective teacher professional development (Little 1993; Guskey 1995; Smylie 1995; Hawley and Valli 1998; Corcoran 1995), resulting in significant rethinking of how professional development is best provided (National Foundation for the Improvement of Education 1996; Darling-Hammond and Ball 1997). This reconceptualization of professional development presents a number of conceptual and technical challenges for cost studies, including methods for assigning costs to professional development activities that are integrated into the instructional day and/or more informal interactions among teachers (Rice 1999).

It is likely that the desire of policymakers and researchers to obtain information regarding appropriate investment levels in teacher professional development will continue to grow. Consequently, research in this area will need to focus increased attention on the development of new conceptual frameworks and cost analyses which can appropriately consider the full array of delivery systems and approaches to providing teacher professional development.

In our analyses, therefore, we decided to focus exclusively on one expenditure account that we believe encapsulates general spending on professional development.

Data and Methods

Our two national studies of district-level spending on professional development relied heavily on two publicly available data sets: (1) the U.S. Bureau of the Census' *Survey of Local Government Finances: School District Finances* (F-33), a school district fiscal report compiled by the U.S. Bureau of the Census; and (2) the Common Core of Data, which is compiled by the National Center for Education Statistics (NCES) and which includes detailed organizational and demographic data on U.S. school districts. We focused on two universe years of data: 1991–92 and 1994–95.

The F-33 report includes general revenue and expenditure data along popular fiscal categories like revenue from property taxes, sales taxes, and a range of user charges, as well as current spending on instruction, salaries and capital accounts. We had hoped to make use of the detailed revenue data at the federal and state levels, but quickly discovered that it is difficult to separate out parts of the various revenue streams by school district. One may identify programs like Title I and Eisenhower grants, but it is not possible to identify the portion of each grant that is earmarked for professional development activities. In our analyses, therefore, we decided to focus exclusively on one expenditure account that we believe encapsulates general spending on professional development. In particular, we have focused on a data element called: "Instructional Staff Support," which is defined by the U.S. Bureau of the Census to include:

> Supervision of instruction service improvements, curriculum development, instructional staff training, and media, library, audiovisual, television, and computer-assisted instruction services.

Ideally, we would have liked to disentangle this expenditure item and separate spending by professional development from elements of instructional staff support. Our goal was to be as precise as possible in the measurement and analysis of the investments of professional development resources into school district staffs. The broadness of the measure is a shortcoming and warrants caution when making comparisons with other more narrowly focused indicators.

As we worked more closely with the F-33 data and developed our analytical strategy, we encountered a significant number of data issues. The U.S. Bureau of the Census as well as the NCES face formidable problems as they seek to gather information in a comparable form from each of the states and territories in the nation. There is a tendency for inconsistencies and surprises to enter the data, and analysts must be on guard for unexpected results that require special interpretation. Moreover, the collection is so vast that it is unreasonable to expect the collectors to understand and anticipate all of the questions that may be raised by a given researcher with a particular set of interests. The best way to improve these collections is for them to be put to use and for the researchers to report back on their experiences at making sense of the data. Some important efforts along these lines have been made, perhaps most notably by O'Leary and Moskowitz (1995). We seek to contribute to this tradi-

tion, and for this reason devote a significant amount of attention in our paper to data issues and our responses. We do this in the hope of stimulating the interest of other researchers in making use of an important data collection.

Using National Survey Data for Local Analysis

There are two national data sets commonly used for fiscal studies at the school district level. *The Annual Survey of Local Government Finances: Public Elementary/Secondary Education Finance Data*, conducted by the U.S. Bureau of the Census, offers the richest source of school district level fiscal data. This survey reports detailed revenue and expenditure information by function for more than 16,000 U.S. school districts. The F-33, as it is commonly called, is one of a battery of local government fiscal surveys. The *Annual Survey of Local Governments* conducted by the U.S. Bureau of the Census accounts for the fiscal environments of roughly 85,000 counties, cities, townships, special districts, and school districts, the five main local government types categorized by the U.S. Bureau of the Census. A nationally representative sample frame is used to determine fiscal conditions around the United States. As part of the Census of Governments, the U.S. Bureau of the Census conducts a complete fiscal census of all local governments in years ending with 2 and 7, which includes school districts.[3] The F-33 is slightly unique in that universe years have been collected in years when only a sample frame is used to survey other local governments. Even when sampling is used to complete the F-33, most states are fully reported and sampling is employed in a minority of states. In 1992, the data content for the F-33 was significantly expanded. Universe years for the F-33 are currently available for 1994–95 and 1995–96.

Of the nine items identified under expenditures for school district support services in the F-33, one variable identifies total expenditures for instructional staff support.

The U.S. Department of Education also collects some fiscal data as part of the annual Common Core of Data (CCD) surveys. The CCD offers a comprehensive database on all schools and school districts regarding contact information, staffing counts, enrollment counts and basic fiscal conditions. The data are not reported via a survey, like the F-33, but rather through contact with state education departments. The CCD serves as the main database for selecting national sample frames for smaller, more detailed surveys conducted by the federal government. While not widely used, some fiscal data covering revenue and expenditures of federal dollars at the school district level are also collected under the General Education Provisions Act, known as the GEPA data files. Due to diligence on the part of survey designers, the F-33 and CCD database files may be joined together through one of three unique school district identifier codes: The NCES ID Code, the U.S. Bureau of the Census Local Agency Code, and the State Government ID Code.

Fiscal Data on Staff Development Expenditures

The F-33 contains district level information about what the U.S. Bureau of the Census calls "instructional staff support services." Of the nine items identified under expenditures for school district support services in the F-33, one variable identifies total expenditures for instructional staff support (variable name = E07). As we noted earlier, "instructional staff support" is defined as those expenditures that include supervision of instruction service improvements, curriculum development, instructional staff training, and media, library, audiovisual, television, and computer assisted instruction services. According to definitions in the NCES *Financial Accounting for Local and State School Systems, 1990* (Fowler 1997), instructional staff support is composed of two main categories: improvement of instruction services and educational media services. The former clearly encapsulates an intuitive conception of expenditures for teacher support services or staff development. Items for this section include:

- Activities concerned with directing, managing, and supervising the improvement of instructional services.
- Activities that assist instructors in designing curriculum, using special curriculum materials, and

[3] This is typical of most fiscal surveys of local government. But, universe data is available for the F-33 for 1994–95, 1995–96, and 1996–97.

learning of techniques to stimulate and motivate students.

- Activities that involve improving the occupational health or professional training of instructional staff, including expenditures for workshops, demonstrations, school visits courses for college credit, sabbatical leave, and travel leaves.

The second major component, educational media services, includes expenditures for activities related to managing and directing educational media, school library services, and audiovisual services. The intent of this component is to capture costs associated with use and preparation of those devices, content materials, methods or experiences used for teaching and learning purposes. The emphasis here is not on training of instructional staff to use the library services or other audiovisual materials, per se, but rather on the general personnel and materials costs involved with preparing audiovisual and other media for use by staff and students. Textbooks are not intended to be charged to this component.

We recognize the fact that interpretations of traditional staff improvement spending (workshops, tuition, in-service training, etc.) are clouded when items like media services are included in a variable such as instructional staff support. However, the variable might also underestimate true staff improvement spending because it does not account for the time costs involved with participation in instructional staff support training. For example, if a teacher attends a day-long training seminar during regular school hours then the provision of a substitute teacher is an added cost on top of travel, registration and other material costs of that seminar. Normally the salaried teacher will also be paid for attending that seminar. A broader definition of professional development costs therefore would need to include these additional salary and benefit payouts to the teacher, among other less obvious costs. Emerging work from the Consortium for Policy Research and Education (CPRE) at the University of Wisconsin utilizes a broader definition of professional development costs like this and has found that the traditionally unobserved time costs greatly exceed what is typically listed as spending on professional development. This example highlights the difficulty in accounting for true investments in professional development and associated expenditures. However, by emphasizing cross-sectional comparisons and longitudinal analyses, we believe we can use the F-33 variable to provide useful insight into the general patterns of school district spending on instructional staff development expenditures.

…by emphasizing cross-sectional comparisons and longitudinal analyses, we believe we can use the F-33 variable to provide useful insight into the general patterns of school district spending on instructional staff development expenditures.

Merging the F-33 and CCD for Database Creation

We utilized the F-33 database cleaning protocols developed by O'Leary and Moskowitz (1995) in order to identify standard operating school districts from other administrative units surveyed in the F-33. In the protocols, the authors summarize the steps employed by three major school finance research groups to clean and maintain a consistent database for school finance research (see table 1). After close inspection, however, we found that only four of the seven recommendations by O'Leary and Moskowitz were useful and found the need to add two new steps to the process.

Although the recommendations of O'Leary and Moskowitz are very useful for winnowing out aberrant school districts, several steps proved questionable in our efforts. O'Leary and Moskowitz note enrollment discrepancies between counts in the F-33 and CCD. We found, as they did in 1995, that several cases with egregious enrollment discrepancies were due to miscoding of school districts with the same names in the same states. No treatment was suggested for this enrollment issue. However, when blending the two data sets O'Leary and Moskowitz recommend merely replacing a missing enrollment record with an available enrollment count from the other survey. Given that this would introduce an uncertain element of bias into the study, we skipped this step entirely. If enrollment counts were missing after regular winnowing by district types, then the entire record was also removed. All per-pupil statistics reported in this paper are based on the F-33 enrollment counts. Second, removal of records where Individualized Education Program (IEP) counts exceeded 50 percent of total enrollment had unintended consequences. Since no IEP counts were recorded for Kentucky in 1991–92 and again in 1994–95, as well as Ohio, Oklahoma, Pennsyl-

Table 1.—Steps taken to join Common Core of Data (CCD) and F-33 datafiles: Methods compared

Steps	O'Leary and Moskowitz (1995)	Killeen, Monk, and Plecki (1999)
1	Merge CCD and F-33 to replace missing enrollments.	Skipped
2	Purge out special or non-operating districts based on the F-33 district types.	Adopted
3	Purge out non-operating districts based on CCD district type codes.	Adopted
4	Purge out districts based on F-33 and CCD district level and grade-span codes.	Adopted
5	Purge districts with zero enrollments or zero revenues and expenditures.	Adopted
6	Purge districts with VOC, TECH, SPEC or AGRIC, in their names.	Skipped
7	Purge districts with greater than 50 percent of their enrollment classified as special education.	Skipped
8		Removed aberrant States from certain years, rather than impute for missing values.
9		Adjusted expenditures by Chambers (1998) Geographic Cost of Education Index

SOURCE: Adapted from O'Leary, Michael and Moskowitz, Jay. 1995. "Proposed 'Good Practices' for Creating Data bases from the F-33 and CCD for School Finance Analyses." In William J. Fowler, Jr. (Ed.), *Selected Papers in School Finance, 1995*. Available online at http://nces.ed.gov/pubs97/97536-5.html.

vania, and Virginia for 1991–92, removal of records based on the recommended criteria would remove the entire state from consideration.[4] Third, the text search for special school districts not immediately winnowed by the F-33 and CCD district types proved onerous and unproductive. In both the 1991–92 and 1994–95 data files, only a handful of records were found that met their criteria. This step was also abandoned. These steps, including removal of States with unreported data for instructional staff support (discussed below), reduced the original number of records in the F-33 by between 20–25 percent for both survey years.

Dealing with the F-33: Handling Missing Records for Instructional Staff Support

Even with the basic database development steps, our research still required handling of those records with missing data for instructional staff support. Unfortunately, the F-33 does not differentiate a missing value for that of a value equal to zero. No flags indicate whether a school district spends zero on teacher professional development or failed to report any spending for this item. During the F-33 universe years, approximately one-third of all states report some level of missing values for the instructional staff support. Our research identified those states high missing values relative to the total number of school districts in the modified data set. States with missing values above 15 percent were identified.

Imputation of those missing records is desirable but premature in our research. One main purpose of our work with the Center for the Study of Teaching and Policy is to understand the empirical and institutional foundation for expenditures on staff development activities. By estimating spending on instructional staff support, prior to a full understanding of what goes into this expenditure item, is hasty and unnecessary. The empirical research is generally sound and will greatly expand our understanding of the conditions under which school districts expend resources for staff development. However, without a rich contextual database for each school district, imputation for missing records through statistical inference would add little to our understanding. We feel as though

[4] According to CCD file documentation from 1991–92, no IEP counts were reported for Guam, Kentucky, Ohio, Oklahoma, Pennsylvania, Puerto Rico, or Virginia. Louisiana counts included only students in self-contained classrooms. New Hampshire figures declined from the previous year because a reporting error was corrected. Sizable changes from 1990–91 are generally associated with an increase in the number of agencies for which IEP counts were reported (U.S. Department of Education 1998).

we can generate a valid empirical understanding of the magnitude of instructional staff development spending in school districts as well as general trends, but know very little about the contextual circumstances that determine that spending. Removal of aberrant states was therefore more acceptable than imputation. As such, for our cross-sectional work based upon the 1994–95 data set, we removed California, Montana, Nebraska, Nevada, and North Dakota from the analyses.

The longitudinal analysis also revealed additional data problems. Several states, such as Tennessee and New Jersey, displayed implausible growth rates for per-pupil spending on instructional staff support. With these findings we re-examined our database methodology but found it to be sound on two levels. First, neither state experienced dramatic enrollment change over the study period. Holding expenditures for instructional staff support constant, enrollment growth would be expected to decrease per-pupil expenditures. The reverse is also true. But neither state exhibited such enrollment changes. Second, neither New Jersey nor Tennessee reported significant (greater than 15 percent) missing records for our target variable. In the course of our research, we found that school districts in 11 states reported incomplete data to the U.S. Bureau of the Census. The five states from 1994–95 are listed above. In 1991–92, those states were Alaska, Arizona, Massachusetts, Maine, New Jersey, and Tennessee. Although the U.S. Bureau of the Census imputed values for many of these districts,[5] we still found the growth statistics to be implausible and therefore excluded all 11 states from our longitudinal analyses.

Chambers' …GCEI was used to adjust for regional differences in instructional staff development expenditures that stem from differences in the cost of key inputs into the educational process.

Geographic Cost Index[6]

Comparison of school districts across rural and urban continuums, as well as region, requires standardization of educational costs. For school districts these differences arise from several sources, including variation in the salaries that must be paid to hire and retain teachers, as well as variation in the extent and nature of the educational services being delivered. Controlling for costs also affords a proximate measure by which to adjust expenditures by geography (Chambers 1998, xi).

Chambers' 1998 release of the Geographic Cost of Education Index (GCEI) was used to adjust for regional differences in instructional staff development expenditures that stem from differences in the cost of key inputs into the educational process. Chambers used a hedonic wage model to predict cost differences for each U.S. school district. The GCEI relies on three main input categories: certified school personnel, noncertified school personnel, and nonpersonnel inputs like supplies, furnishings, utilities, and contract expenditures (Chambers 1998, 7). The GCEI is available for the years 1990–91 and 1993–94. These index years were used to adjust our databases for the years 1991–92 and 1994–95, respectively. The implication of this mismatch is truly unknown, though likely to be small for two reasons. Chambers' (1998) research indicates an extremely high correlation of GCEI indices over a period of 6 years, indicating that GCEI estimates for 1 year are a suitable estimate for another year. Second, local economies on the whole tend to shift in period fashion, rather than abruptly. Therefore, changes on a year-to-year basis will likely be small and of minimal impact on school input costs. These two points are assumptions and limitations with our database creation. As more specific cost of education indices becomes available, we will readjust our database. Chambers (1998) does note that the GCEI tends to minimize differences between school districts in terms of expenditures, which would mean measurement of expenditure inequality in our database will likely be smaller than in reality.

[5] The U.S. Bureau of the Census indicated that individual records in these states were estimated one of two ways. If a minority of school districts in the state could be accurately estimated based off of share ratios from other districts in the state, those ratios were used to impute the missing records. Alternatively, some records in states were imputed using national share ratios, if that state was representative of the entire nation. Missing records were not imputed for some states, mainly the ones we identified earlier, because of uniqueness in their structure, i.e., extremely small, rural districts in Montana. Sharon Meade of the Governments Division, U.S. Bureau of the Census, described these database limitations to Kieran Killeen (7/99).

[6] This section borrows heavily from our working paper, recently submitted to the *Journal of Education Finance* (see Killeen, Monk, and Plecki 1999).

Choosing Reporting Statistics and Interpreting Results

Our initial research on professional development spending focused primarily on the differences in staff improvement expenditures across place and time. Comparison of resources by place requires standardization by population size. As per conventions in the school finance literature, our research reports on findings in per-pupil terms and in terms of the share of total general fund expenditures. We recognize that per-pupil expenditures do not intuitively capture an expenditure item that deals almost exclusively with expenditures for teacher development and improvement. A statistic that compares expenditures on staff development standardized by instructional staff size would be both interesting and useful. The Bureau of Labor Statistics reports private sector training expenditures in terms of expenditures per employee. Several issues made it difficult to construct this statistic. It is neither clear in the F-33, nor in the general staff development literature, how staff development dollars are allocated across school district employees. Typically, it is assumed that the vast majority of dollars go towards the teaching staff. However, to what degree administrative aides, administrators, and other specialized school district personnel receive staff development dollars to improve instruction is unclear. In all likelihood, school district allotment formulas for staff improvement dollars may be more similar than different across the United States. Future research may advance our understanding of how personnel categories differentially absorb professional development resources. At this juncture, reporting staff improvement expenditures in per-pupil terms satisfies general weighting criteria, and allows for comparison of resources across space controlling for population size. We also chose to report expenditures as a share of total general expenditures.

On average, in 1994–95, U.S. school districts spent 2.76 percent of total expenditures on instructional staff support.

Findings

The methodology we employed to manipulate the F-33 data served as the base for two sets of analyses of spending on professional development. A brief summary of those findings is presented here.

On average, in 1994–95, U.S. school districts spent 2.76 percent of total expenditures on instructional staff support (see table 2). When reported in per-pupil terms, instructional staff support equates to about $200 per pupil. When summed by state to the national level, 3.32 percent of total expenditures are devoted to instructional staff support. This latter statistic is the weighted average. Table 2 reports both the weighted and simple averages.

We found a reasonable degree of consistency in spending on instructional staff support across all school districts in both per-pupil terms and as a share of total expenditures. We found that most U.S. school districts expend between 2 and 5 percent of their budget on this item. States with school districts exceeding this trend include Kentucky, South Carolina, Tennessee, Virginia, and Florida. School districts across Kentucky, for example, spend on average 8 percent of total expenditures on instructional staff support or more than $500 per pupil, the highest in the nation.

Our analyses also revealed moderate growth in the level of spending on professional development. We found that between 1992 and 1995, spending on instructional staff support grew by 25 percent in per-pupil terms. In terms of the share ratios, we found an 8 percent increase in the average budget share devoted to instructional staff support spending.

The most interesting caveat to our national analysis concerns differences in average spending levels by urbanicity. In preliminary work, we found that urban districts expend more on instructional staff support in per-pupil terms and in terms of total general expenditures. These findings held on a simple three-point urbanicity scale (urban, suburban, and rural). We advanced our analysis by examining expenditure patterns via a seven-point scale readily available in the Common Core of Data (Killeen, Monk, and Plecki 2000).

When districts are coded by urbanicity (see table 3), we found that population density relates to expenditures on instructional staff support; spending increases with density. Districts in large central cities spend 3.43 percent of their budgets on instructional staff support, whereas ru-

Table 2.—State spending on instructional staff support (1994–95): State-by-state comparisons[1]

State[2]	Enrollment	Instructional staff support (ISS in 000's)	Instructional staff support expenditures as a percentage of general expenditures: Weighted average[3]	Instructional staff support expenditures as a percentage of general expenditures: Simple average[4]	Instructional staff support expenditures per pupil: Weighted average	Instructional staff support expenditures per pupil: Simple average
Nation[5]	**37,515,224**	**8,033,816**	**3.32**	**2.76**	**214**	**192**
Top five states ranked by enrollment						
Texas	3,670,007	752,175	3.49	2.57	205	184
New York	2,738,028	469,053	2.00	2.80	171	267
Florida	2,107,514	640,769	4.56	4.46	304	299
Illinois	1,897,161	313,845	2.76	2.07	165	126
Ohio	1,829,761	396,060	3.71	3.00	216	173
Top five states ranked by share of instructional staff support to total expenditures (simple average)						
Kentucky	639,992	311,882	8.14	8.10	487	504
South Carolina	638,548	179,659	4.99	5.20	281	306
Tennessee	870,594	196,846	4.54	4.73	226	237
Virginia	1,058,709	313,716	4.69	4.52	296	294
Florida	2,107,514	640,769	4.56	4.46	304	299

[1] The expenditure data were adjusted using Chambers' 1998 Geographic Cost Index.
[2] The following states were removed from the analysis due to a high proportion of missing values in 1994–95: California, Montana, Nebraska, Nevada, and North Dakota.
[3] The weighted average is calculated as the summation of expenditures per state divided by the total enrollment.
[4] The simple average is calculated as the average value per school district.
[5] The weighted average sums expenditures across the nation divided by the total enrollment.

SOURCE: U.S. Bureau of the Census. 1995. "Survey of Local Government Finances: School District Expenditures (F-33), 1994–95."

ral districts spend 2.46 percent. At $222 per pupil, districts in large central cities spend $40 more than rural districts. It is also interesting to note that as one travels from a center city core, through the suburbs, spending on instructional staff support falls. Spending then climbs in large towns, or places of greater population density.

We speculate that urban districts, over less urban districts, tend to spend more as a function of the higher demand for staff development programming. There are at least two reasons for this expectation. First, with high concentrations of young and inexperienced teachers, urban districts must spend more to train and retain their teaching force. Spending is greater because young teachers participate more frequently in training sessions and change jobs more often. The mobility issue, in particular, causes greater demand for new teacher training. Second, given that urban areas generally contain high poverty populations, federal dollars like Eisenhower Professional Development Program funds tend to flow disproportionately into urban areas.[7] For example, we found evidence that shows that urban school districts do in fact receive more Eisenhower funds, and argue that this could contribute to resources for higher spending.[8]

[7] Federal program funds for professional development activities, especially Eisenhower funds, are also directed to institutions of higher education. Because colleges and universities also concentrate in urban areas, the availability for professional development training opportunities may be higher in urban areas than other places. Urban school districts, therefore, may spend more because more is available.

[8] Ibid.

Table 3.—Instructional staff support expenditures,[1] by school district urbanicity, 1994–95

Urbanicity[2]	School district averages: Instructional staff support expenditures as a percentage of general expenditures	School district averages: Instructional staff support expenditures per pupil
Nationally[3]	**2.76**	**192**
Large central city	3.43	222
Mid-size central city	3.30	215
Urban fringe of large city	2.92	210
Urban fringe of mid-size city	3.03	192
Large town	3.42	208
Small town	3.04	195
Rural	2.46	182

[1] Fiscal data are adjusted using Chambers 1998 Geographic Cost Index.
[2] The urbanicity scale used here is a seven point National Center for Education Statistics (NCES) classification, where:

A. Large city: A central city of a Consolidate Metropolitan Statistical Area (CMSA) or MSA, with the city having a population greater than or equal to 250,000.

B. Mid-size city: A central city of a CMSA or MSA, with the city having a population less than 250,000.

C. Urban fringe of large city: Any incorporated place, Census designated place, or non-place territory within a CMSA or MSA of a large city and defined as urban by the U.S. Bureau of the Census.

D. Urban fringe of mid-size city: Any incorporated place, Census designated place, or non-place territory within a CMSA or MSA of a mid-size city and defined as urban by the U.S. Bureau of the Census.

E. Large town: An incorporated place or Census designated place with population greater than or equal to 25,000 and located outside a CMSA or MSA.

F. Small town: An incorporated place or Census designated place with population less than 25,000 and greater than or equal to 2,500 and located outside a CMSA or MSA.

G. Rural: Any incorporated place, Census designated place, or non-place territory designated as rural by the U.S. Bureau of the Census.

[3] This statistic represents a simple average of all school districts at the national level, then along the seven-point urbanicity scale.

SOURCE: U.S. Bureau of the Census. "Survey of Local Government Finances: School District Expenditures (F-33), 1991–92;" U.S. Bureau of the Census. "Survey of Local Government Finances: School District Expenditures (F-33), 1994–95;" and U.S. Department of Education, National Center for Education Statistics, Common Core of Data, 10-year longitudinal file.

Conclusion and Suggestions for NCES

In our two studies of professional development expenditures, the F-33 has proven to be a useful starting point for estimating local spending patterns within the national context. It is interesting to note that the magnitudes we find for the instructional staff support variable, as a proxy for total spending on professional development, are reasonably consistent with the array of findings from the case study research on this topic. In particular, the case study research on professional development spending, where individual budget records are analyzed on a case-by-case basis, researchers have found budget share ratios range between 1.8 to 3.0 percent (Little et al. 1987; Miller, Lord, and Dorney 1994; and Elmore 1997).

In this effort to analyze district level professional development spending patterns across the United States, the F-33 database has proven to be quite useful. Modifications to the database, including record cleaning techniques, are not difficult to administer. Researchers should continue to utilize the basic database cleaning techniques outlined by O'Leary and Moskowitz (1995) in order to standardize comparative studies based upon fiscal analyses of the F-33. We also feel researchers should continue to connect the F-33 with cost adjustment indexes such as those produced by Chambers (1998), as these indexes minimize the cost of education differences when comparing district fiscal patterns across the nation. In addition, the F-33 is easily linked with other NCES data sets through unique record identifiers. We found that these

unique identifiers made district level fiscal records quite portable and easy to join with variables from the Common Core of Data. Our work with existing national data sets also highlight some inconsistencies in the conventions used to discuss total expenditures on teacher professional development.

As noted earlier, a number of cost accounting issues continue to cloud clear estimates of the total expenditures made on teacher professional development. For example, greater attention needs to be paid to the amount of time teachers and administrators are spending participating in training activities. Narrow descriptions of traditional professional development expenditures seem to avoid the importance of time costs, or those unaccounted salary and benefit costs of having teachers and staff participate in professional development training activities. Narrow descriptions also fail to include the importance of salary credits, the dollar amount that districts pay to staff over their careers for participating in training activities. There is no one right way to account for the total expenditures on teacher professional development, but clearly standards are needed. The NCES remains uniquely positioned to enhance existing and future databases to provide more consistent information about total spending on professional development.

…a number of cost accounting issues continue to cloud clear estimates of the total expenditures made on teacher professional development.

By providing new accounting standards for professional development, the NCES could foster greater consistency and agreement in the analysis of effective professional development training investments. There are essentially three areas where standards could substantively improve the quality of data and therefore enhance research opportunities. First, there is an absence of clear information on how professional development activities are funded. Specifically, greater information is needed on the share of federal, state and local fund sources, as well as the programmatic basis for the fund sources. Great emphasis, for example, is placed on federal Eisenhower Professional Development funds for improving the quality of teachers, but little is known about the collective effect of all federal program dollars for this purpose. Second, new standards could help focus attention on what is actually purchased with professional development resources. Limitations with existing national databases do not allow us to separate out professional development expenses by personnel status (teachers, administrators, or staff), or by type of expense (salary credit, travel, tuition reimbursement, registration, etc.). This limitation hinders the opportunity to focus attention on training teachers, as well as the opportunity to understand what are the major and significant costs of that training. The third benefit of new standards speaks to the need to tie investments in teacher training to traditional measures of equity as well as outcomes. Very little is known about how professional development dollars are distributed, whether they are spread evenly across and within districts, or tend to concentrate in particular areas, such as places of high poverty and teacher shortages. New standards could also foster the opportunity to connect the investments in teacher training with student outcomes. These three elements represent target areas by which to measure the effectiveness of new accounting standards for professional development programming.

Translating these standards into new data collection efforts would provide a substantial contribution to empirical research on teaching and learning. For example, with enhanced data on what is actually purchased with professional development resources, researchers could begin to explore what specific investments contribute to gains in overall student performance, the performance gains of low-income students, or even the types of investments that move poorly performing students to greater achievement. A national sample of school district finances, via an instrument that uses the new standards, could provide this data. Perhaps the closest opportunity to build a new data set exists in refinements to existing national surveys, such as the Schools and Staffing Survey (SASS), the National Educational Longitudinal Study (NELS), the Early Childhood Longitudinal Study (ECLS), or other such surveys conducted by the National Center for Education Statistics. At a minimum, the sample frames for these national surveys could be adopted and a fiscal survey could be conducted. Approached in this fashion, the blending of new standards for data collection on teacher professional development and linking to existing national databases would significantly improve research on effective teaching and learning.

References

Chambers, J. 1998. *Working Paper: Geographic Variations in Public School's Costs.* Washington, DC: U.S. Department of Education, National Center for Education Statistics (NCES Working Paper 98–04).

Corcoran, T.B. 1995. *Helping Teachers Teach Well: Transforming Professional Development.* New Brunswick, NJ: Consortium for Policy Research in Education Policy Briefs.

Darling-Hammond, L. and Ball, D. 1997. *Teaching to High Standards: What Policymakers Should Know and Be Able to Do*. A report to the National Education Goals Panel (http://www.negp.gov/Reports/highst.htm).

Education Commission of the States. 1997. *Investment in Teacher Professional Development: A Look at 16 Districts*. Denver, CO: Author.

Elmore, R. 1997. *Investing in Teacher Learning: Staff Development and Instructional Improvement in Community School District #2, New York City*. Washington, DC: National Commission on Teaching and America's Future, Consortium for Policy Research in Education.

Fowler, Jr., W.J. 1997. *Financial Accounting for Local and State School Systems, 1990.* Washington, DC: U.S. Department of Education, National Center for Education Statistics (NCES 97–096R).

Guskey, T. 1995. "Professional Development in Education." In T.R. Guskey and M. Huberman (Eds.), *Professional Development in Education*. New York, NY: Teachers College Press. 114–134.

Hawley, W.D. and Valli, L. 1998. "The Essentials of Effective Professional Development: A New Consensus." In L. Darling-Hammond (Ed.), *The Heart of the Matter*. San Francisco, CA: Jossey-Bass.

Killeen, K., Monk, D., and Plecki M. 2000. "Spending on Instructional Staff Support Among Big City School Districts: Why are Urban Districts Spending at Such High Levels?" *Educational Considerations*. 28(1).

Killeen, K., Monk, D., and Plecki M. 1999. "School District Spending on Professional Development: Insights from National Data." A Working Paper of the Center for the Study of Teaching and Policy, University of Washington.

Little, J.W. 1993. "Teachers' Professional Development in a Climate of Educational Reform." *Educational Evaluation and Policy Analysis.* 15(2): 129–151.

Little, J.W. 1989. "District Policy Choices and Teacher's Professional Development Opportunities." *Educational Evaluation and Policy Analysis.* 11(2): 165–179.

Little, J.W., Gerritz, W.H., Stern, D.H., Guthrie, J.W., Kirst, M.W., and Marsh, D.D. 1987. *Staff Development in California.* San Francisco, CA: Policy Analysis for California Education (PACE) and Far West Laboratory for Educational Research and Development (Policy paper #PC87-12-15, CPEC).

Lytle, J.H. 1983. "Investment Options for Inservice Teacher Training. *Journal of Teacher Education* 34(1): 28–31.

Miller, B., Lord, B., and Dorney J. 1994. *Staff Development for Teachers: A Study of Configurations and Costs in Four Districts.* Newtonville, MA: Education Development Center.

Moore, D. and Hyde, A. 1981. *Making Sense of Staff Development: An Analysis of Staff Development Programs and Their Costs in Three Urban School Districts*. Chicago: Designs for Change.

National Foundation for the Improvement of Education. 1996. *Teachers Take Charge of Their Learning: Transforming Professional Development for Student Success*. Washington, DC: Author.

O'Leary, Michael and Moskowitz, Jay. 1995. "Proposed 'Good Practices' for Creating Data Bases from the F-33 and CCD for School Finance Analyses." In William J. Fowler, Jr. (Ed.), *Selected Papers in School Finance, 1995*. Available online at http://nces.ed.gov/pubs97/97536-5.html.

Rice, J.K. 1999. *Recent Trends in the Theory and Practice of Teacher Professional Development: Implications for Cost.* Seattle, WA: Paper prepared for the annual conference of the American Education Finance Association.

Ross, R. 1994. *Effective Teacher Development Through Salary Incentives (An Exploratory Analysis).* RAND, Institute on Education and Training.

Smylie, M.A. 1995. "Teacher Learning in the Workplace: Implications for School Reform." In T.R. Guskey and M. Huberman (Eds.), *Professional Development in Education.* New York, NY: Teachers College Press. 92–113.

Stern, D.S., Gerritz, W., and Little, J.W. 1989. "Making the Most of the District's Two (Or Five) Cents: Accounting for Investment in Teachers' Professional Development." *Journal of Education Finance.* 14(4): 19–26.

U.S. Bureau of the Census. "Survey of Local Government Finances: School District Finances (F-33)," Available online at http://www.census.gov/govs/www/school.html.

U.S. Department of Education, National Center for Education Statistics. Common Core of Data: Public Education Agency Universe, 1991–92 [Computer file]. Conducted by the U.S. Department of Commerce, U.S. Bureau of the Census. ICPSR version. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor], 1998.

Making Money Matter: Financing America's Schools

Helen F. Ladd

Duke University

Janet S. Hansen

Committee for Economic Development

About the Authors

Helen F. Ladd is professor of public policy studies and economics at the Terry Sanford Institute of Public Policy at Duke University. Her current research focuses on education policy, including performance-based approaches to reforming schools, school choice, education finance, and teacher quality. She is editor of *Holding Schools Accountable: Performance-Based Reform in Education* and co-author (with Edward B. Fiske) of *When Schools Compete: A Cautionary Tale.* Her current research projects focus on teacher quality and student achievement in the United States and financing education in South Africa.

In addition, she has published extensively in the areas of property taxation, state economic development, and the fiscal problems of U.S. cities. She is the author of *Local Government Land Use and Tax Policies in the United States: Understanding the Links* and (with John Yinger) *America's Ailing Cities: Fiscal Health and the Design of Urban Policy.* She has been a visiting scholar at the Federal Reserve Bank of Boston, a senior fellow at the Lincoln Institute of Land Policy, a visiting fellow at the Brookings Institution, and a Fulbright lecturer/researcher in New Zealand and in South Africa. She has a Ph.D. in economics from Harvard University.

Janet S. Hansen is currently Vice President and Director of Education Studies at the Committee for Economic Development in Washington, DC. As senior program officer at the National Research Council (NRC) from 1991–99, she served as staff director for the Committee on Education Finance. While at the NRC, she managed several other projects related to education and training, international comparative studies in education, and civilian aviation careers. Prior to joining the NRC staff, she was director for policy analysis at the College Board. She has written and lectured widely on issues relating to higher education finance, federal and state student assistance programs, and how families pay for college. She also served as director for continuing education and associate provost at Claremont College and as assistant dean of the college at Princeton University. She has a Ph.D. in public and international affairs from Princeton University.

Making Money Matter: Financing America's Schools

Helen F. Ladd

Duke University

Janet S. Hansen

Committee for Economic Development

Acknowledgments

The following executive summary has been extracted and reprinted with permission from publication of the same title, originally published by the National Academy Press (NAP) for the National Research Council's Committee on Education Finance. The full publication can be found at the NAP Web Site: http://www.nap.edu/catalog/9606.html.

Executive Summary

A national desire to ensure that all children learn and achieve to high standards now poses fundamental challenges to almost every facet of business as usual in American education. Policymakers and educators are searching for better ways to provide today's schoolchildren with the knowledge and skills they will need to function effectively as citizens and workers in a future society that promises to be increasingly complex and globally interconnected. A key component of this quest involves school finance and decisions about how the $300 billion the United States spends annually on public elementary and secondary education can most effectively be raised and used.

A new emphasis on raising achievement for all students poses an important but daunting challenge for Policymakers: how to harness the education finance system to this objective. This challenge is important because it aims to link finance directly to the purposes of education. It is daunting because making money matter in this way means that school finance decisions must become intertwined with an unprecedented ambition for the nation's schools: never before has the nation set for itself the goal of educating all children to high standards.

This report argues that money can and must be made to matter more than in the past if the nation is to reach its ambitious goal of improving achievement for all students. There are, however, no easy solutions to this challenge, because values are in conflict, conditions vary widely from place to place, and knowledge about the link between resources and learning is incomplete. Moreover, without societal attention to wider inequalities in social and economic opportunities, it is unrealistic to expect that schools alone, no matter how much money they receive or how well they use it, will be able to overcome serious disadvantages that affect the capacity of many children to gain full benefit from what education has to offer.

Taking full account of conflicting values, wide variation in educational contexts, and strengths and limitations of existing knowledge, the Committee on Education Finance concludes that money can and should be used more effectively than it traditionally has been to make a difference in U.S. schools. To promote the achievement of a fair and productive educational system, finance decisions should be explicitly aligned with broad educational goals. In the past, finance policy focused primarily on availability of revenues or disparities in spending, and decisions were made independently of efforts to improve the educational system's performance. Although school finance policy must not ignore the continuing facts of revenue needs and spending disparities, it also should be a key component of education strategies designed to foster higher levels of learning for all students and to reduce the nexus between student achievement and family background.

To this end, the emerging concept of funding adequacy, which moves beyond the more traditional concepts of finance equity to focus attention on the sufficiency of funding for desired educational outcomes, is an important step. The concept of adequacy is useful because it shifts the focus of finance policy from revenue inputs to spending and educational outcomes and forces discussion of how much money is needed to achieve what ends. It also could drive the education system to become more productive by focusing attention on the relationship between resources and outcomes.

Making money matter more requires more than adequate funding.

Applying an adequacy standard to school finance is at present an art, not a science. Misuse of the concept can be minimized if adequacy-based policies are implemented with appropriate recognition of the need for policy judgments and of the incomplete knowledge about the costs of an adequate education. Efforts to define and measure adequate funding are in their infancy. A number of technical challenges remain, including the determination of how much more it costs to educate children from disadvantaged backgrounds than those from more privileged circumstances. Beyond these, some fundamental questions about educational adequacy (such as how broad and how high the standards should be) are ultimately value judgments and are not strictly technical or mechanical issues. A key danger is that political pressures may result in specifying adequacy at so low a level as to trivialize the concept as a meaningful criterion in setting finance policy, or at so high a level that it encourages unnecessary spending. Another is that Policymakers will fail to account for the higher costs of educating disadvantaged students.

Making money matter more requires more than adequate funding. It also requires additional finance strategies, such as investing in the capacity of the education system, altering incentives to ensure that performance counts, and empowering schools or parents or both to make decisions about the uses of public funds. For money to matter more, it must be used in ways that ensure that schools will have the capacity to teach all students to higher standards as well as the incentive to do so. Policy options involve choices among individual finance strategies and combinations of strategies; policy decisions will depend partially on philosophical outlook but can also be informed by careful attention to evidence from research and practice. Attention to context is important as well, as educational and political conditions diverge widely from place to place and individual policy options will often vary in effectiveness depending on local circumstances.

Educational challenges facing districts and schools serving concentrations of disadvantaged students are particularly intense, and social science research provides few definitive answers about how to improve educational outcomes for these youngsters. While pockets of poverty and disadvantage can be found in all types of communities, the perceived crisis in urban education is especially worrisome. Ongoing reform efforts should be encouraged and evaluated for effectiveness. At the same time, systematic inquiry is needed into a range of more comprehensive and aggressive reforms in urban schools. Piecemeal reform efforts in the past have not generated clear gains in achievement, and generations of at-risk schoolchildren have remained poorly served by public education. Because the benefits of systematic inquiry will extend beyond any one district or state, the federal government should bear primary responsibility for initiating

and evaluating bold strategies for improving education for at-risk students.

Improving the American system of education finance is complicated by deeply rooted differences in values about education, the role of parents in guiding the development of their children, and the role of individuals and governments in a democratic society. In addition, there are serious shortcomings in knowledge about exactly how to improve learning for all students. Education policy cannot ignore these facts. Instead, the challenges are to balance differing values in a thoughtful and informed manner and continuously to pursue bold, systematic, and rigorous inquiry to improve understanding about how to make money matter more in achieving educational goals. The committee is convinced that these challenges can be met and that the nation can improve the way it raises and spends money so that finance decisions contribute more directly to making American education fair and effective.

The Committee's Charge and Approach

The Committee on Education Finance was established under a congressional mandate to the U.S. Department of Education to contract with the National Academy of Sciences for a study of school finance. In fleshing out the brief mandate assigned from Congress, the department charged the committee to evaluate the theory and practice of financing elementary and secondary education by federal, state, and local governments in the United States. The key question posed to the committee was: *How can education finance systems be designed to ensure that all students achieve high levels of learning and that education funds are raised and used in the most efficient and effective manner possible?* In carrying out its study, the committee was further charged to give particular attention to issues of educational equity, adequacy, and productivity.

Finance policy and practice, especially now that they are being linked to the nation's highest ambitions for schools, touch on virtually all facets of education.

The committee translated these key questions into three goals for education finance systems. This translation provided objectives against which to evaluate the performance of existing arrangements and the likely effects of proposed changes:

Goal 1: Education finance systems should facilitate a substantially higher level of achievement for all students, while using resources in a cost-efficient manner.

Goal 2: Education finance systems should facilitate efforts to break the nexus between student background characteristics and student achievement.

Goal 3: Education finance systems should generate revenue in a fair and efficient manner.

Finance policy and practice, especially now that they are being linked to the nation's highest ambitions for schools, touch on virtually all facets of education. Inevitably, therefore, finance is controversial; education policy is one of the most contentious items on the public policy agenda because it is deeply enmeshed in competing public values. Widespread support for equality of educational opportunity masks disagreement over the extent to which high levels of fiscal equality among students or between school districts is required and over the extent to which it is appropriate for parents to spend some of their resources to benefit their own children in preference to others. The division of powers in U.S. government and a traditional emphasis on local control make changes in the dispersion of responsibilities for raising and spending education dollars difficult and slow. Americans' deep belief in the value of efficiency becomes complicated to act on when it encounters limited knowledge about what efficient solutions are in education, disagreements about what the ends of education should be, and belief that the educational system should be democratically governed and responsive to a variety of local, state, and national needs and views. It is thus hard for schools to be both democratic institutions and to have the focused and durable goals that are viewed by some as necessary for an efficient system.

Education policy in general and finance policy more specifically raise difficult questions that require both moral wisdom and empirical research. Experts, such as the members of the Committee on Education Finance, can contribute to policy making by examining evidence and by rationally and objectively clarifying the values and objectives at stake. They cannot resolve all disagreements, but

they can render some views more reasonable and others less so.

The committee's inquiry into education finance takes place against the backdrop of a highly decentralized and diverse system of U.S. education that makes description and generalization difficult. The existing finance system is broadly characterized by delegation of significant responsibility for education to the local level, by an average division of funding responsibilities roughly even between state and local governments (with the federal government providing only about 7 percent of education revenues available to schools), and by great variation from place to place in the funds available for education and the level of government that provides them. Education is not mentioned in the federal Constitution and therefore has been viewed as a power reserved to the states, most of whose constitutions specify the provision of education as a key state obligation.

Another backdrop for the committee's deliberations is its assessment of the current condition of education as it relates to the three goals. Regarding goal 1—promoting higher achievement for all students—and goal 2—reducing the nexus between student achievement and family background—the committee concluded that although schools are not failing as badly as some people charge, they are not sufficiently challenging all students to achieve high levels of learning and are poorly serving many of the nation's most disadvantaged children. The continuing correlation between measures of student achievement and student background characteristics, such as ethnic status and household income, looms ever more serious as global economic changes have increasingly tied the economic well-being of individuals to their educational attainment and achievement. Particularly troublesome is the perceived crisis in education in many big-city school systems, a condition that has concerned Policymakers since the 1960s but has been too often stubbornly resistant to improvement.

Regarding goal 3—raising revenue fairly and efficiently—the United States is unique in its heavy reliance on revenue raising by local school districts, the extensive use of the local property tax, and the small federal role. Despite significant amounts of state financial assistance to local school districts, spending levels vary greatly among districts within states and also across states, a situation that many people believe is unfair. Moreover, the local property tax is not always administered equitably and may generate a greater burden on taxpayers with low income than on those with high income. Efforts to increase fairness, however, must be balanced by sensitivity to possible effects on the efficiency with which funds are raised.

Fairness and Productivity in School Finance

Fairness in the distribution of education dollars has long been an objective of school finance reformers, but one that has frequently been thwarted by the political realities of an education system that allocates much of the responsibility for funding and operating schools to local governments. Concern about how funding policies and practices affect the performance of schools is a more recent development, but one that is becoming ever more central to school finance decision making.

*In the aftermath of **Brown v. Board of Education**…the United States awoke from its historical indifference to the problem of unequal educational opportunities and began to address them.*

In the aftermath of *Brown v. Board of Education*, 347 U.S. 483 (1954), the United States awoke from its historical indifference to the problem of unequal educational opportunities and began to address them. Beginning about 1970, the nation entered a notably vigorous period of school finance reform aimed at making the distribution of education dollars more fair. Litigants in a number of states succeeded in having state finance systems overturned in court on the grounds that they violated state constitutional equal protection provisions or education clauses. In the wake of these court decisions, virtually all states, whether under court order or not, substantially changed their finance systems. State and federal governments also created a number of categorical programs directing resources to students with special education needs and to some extent compensating for funding inequities at the local level.

Despite these changes, U.S. education continues to be characterized by large disparities in educational spending. While within-state funding disparities decreased in

some states, especially those subject to court-mandated reform, large disparities persist. Moreover, disparities continue to mirror the economic circumstances of district residents; districts with lower-income residents spend less than districts whose residents have higher incomes. In some districts, this pattern is repeated in school-to-school spending differences. Nationwide, over half of the disparity in district per-pupil spending is the result of differences in spending between states rather than within states.

Particularly in the last decade, the concept of fairness as it applies to school finance has taken on a new emphasis, spawning another round of litigation and reform. The pursuit of fairness has moved beyond a focus on the relative distribution of educational inputs to embrace the idea of educational adequacy as the standard to which school finance systems should be held.

Despite the success of adequacy arguments in several prominent school finance court decisions, there is as yet no consensus on its meaning and only limited understanding about what would be required to achieve it. Adequacy is an evolving concept, and major conceptual and technical challenges remain to be overcome if school finance is to be held to an adequacy standard. Earlier concepts of equity posed similar challenges in their infancy, although over time much progress was made in defining and measuring them. Similar progress may be expected here. In the meantime, awareness of the shortcomings in current understanding of adequacy is important for all who would use the concept in either policy making or in research.

Despite the success of adequacy arguments in several prominent school finance court decisions, there is as yet no consensus on its meaning and only limited understanding about what would be required to achieve it.

In part, efforts to use finance policies to achieve educational adequacy depend centrally on understanding how to translate dollars into student achievement. In fact, however, knowledge about improving productivity in education is weak and contested. The concept itself is elusive and difficult to measure. There is as yet no generally accepted theory to guide finance reforms. Instead multiple theories, each of which is incomplete, compete for attention. Empirical studies seeking to determine the best ways to direct resources to improve school performance have produced inconsistent findings.

Equality of Educational Opportunity, the famous study of the mid-1960s known as the Coleman Report, found that, after family background factors were statistically controlled, school resource variation did not explain differences in student achievement. The Coleman report ushered in decades of productivity research attempting to understand (and perhaps discredit) that counterintuitive result. For many years, the inability of researchers to speak consistently on how to improve schools has frustrated scientists and Policymakers alike. While there is still a great deal of uncertainty about how to make schools better or how to deploy resources effectively, the committee's review of the last several decades of research and policy development on educational productivity makes us more optimistic than our predecessors regarding the prospects for making informed school finance choices. Thirty years' worth of insights have generated a host of ideas about how to use school finance to improve school performance, and researchers have learned to ask better questions and to use improved research designs that yield more trustworthy findings. Knowledge is growing and will continue to grow. One major implication of this fact for school finance is that good policy will reflect both the best knowledge available to date *and* the need to continue experimenting and evolving as new knowledge emerges.

Even while understanding is becoming more sophisticated, knowledge about how to improve educational productivity will always be contingent and tentative, in part because the characteristics and needs of key actors—the students—differ greatly from place to place. Therefore, solutions to the challenge of improving school performance are unlikely ever to apply to all schools and students in all times and places. Policymakers and the public will have to consider evidence and analysis about the strengths and weaknesses of strategies for change as they also weigh differing values about what Americans want their schools to be and to do.

Strategies for Meeting the Goals

Four generic strategies can be used to make money matter more for U.S. schools and to propel the education system in desirable directions:

- Reduce funding inequities and inadequacies;
- Invest more resources (either new or reallocated from other uses) in developing capacity;
- Alter incentives to make performance count (within the existing governance structure); and
- Empower schools and parents to make decisions about the use of public funds (thereby altering governance and management relationships).

Reducing funding inequities and inadequacies includes options such as reducing disparities in funding across schools, districts, or states; ensuring that all schools or districts have funding sufficient to provide an adequate level of education to the students they serve; and raising revenue more fairly without neglecting efficiency. Investing more resources in developing capacity refers not only to the capacity of the formal education system to provide services but also to the capacity of students to learn. Hence, it includes investments in inputs, such as teacher quality and technology, and in programs, such as preschool for disadvantaged students. Altering incentives embraces changes in incentives designed to operate primarily within the existing system of school governance and includes policies such as restructuring teacher salaries, use of school-based incentive programs, and changes to the incentives built into financing formulas for students with special needs. Empowering schools and parents refers to policies that would decentralize significant authority over the use of public funds, to schools in the form of site-based management or charter schools, and to parents in the form of significant additional parental choice over which schools (public and perhaps private as well) their children will attend.

Finance policies ought to reflect the interrelatedness of the various facets of the finance system and the possibility that complementary changes may be required for reform to be successful.

In reality, policymakers do not and should not consider strategies in isolation. Finance policies ought to reflect the interrelatedness of the various facets of the finance system and the possibility that complementary changes may be required for reform to be successful. Indeed, some visions of overall education reform explicitly call for a set of intertwined finance strategies.

Our decision to examine the strategies separately is useful for analytical purposes, but it also reflects the important fact that strategies can be combined in different ways. It is important to emphasize, however, that not all strategies are compatible. For example, a centrally (i.e., state or school district) managed program of investment in capacity would not fit naturally with a program that empowers parents and schools to make decisions about the kind of capacity in which they wish to invest.

For each of the three goals for an education finance system, we evaluate a variety of policy options employing these strategies and weigh the evidence on how effective they are likely to be in helping meet the objectives.

Achieving Goal 1: Promoting Higher Achievement for All Students in a Cost-Efficient Way

- Adequate funding (sufficient funding for efficiently operating schools to generate higher achievement levels) is clearly essential for meeting goal 1. Although we do not know how to identify this level with precision, it is important to try. But providing adequate funding by itself may do little to foster significant improvements in overall student achievement. Thus, while funding adequacy may be a necessary part of any education reform effort—and is likely to be especially crucial for districts or schools serving disproportionate numbers of disadvantaged students—it is at most part of an overall program for increasing student achievement in a cost-efficient way.
- Teaching all students to higher standards makes unprecedented demands on teachers and requires changes in traditional approaches to teacher training and retraining. In addition to nonfinance policies for investing in the capacity of teachers (e.g., reforming teacher preparation and licensing), finance options might include raising teacher salaries and investing in the professional development of teachers once they are on the job. Given schools' need to hire 2 million new teachers over the coming decade, raising salaries—especially for new hires—may be needed to ensure sufficient numbers of qualified people in classrooms. Professional development that is aligned with curriculum re-

form and teaching objectives offers the promise of changing teaching practice in ways likely to improve student performance. But neither approach is likely to be effective in achieving goal 1 unless it is aligned with appropriate incentives throughout the education system to make performance count.

- Altering incentives responds to the fact that the school finance system historically has operated almost in isolation from educational performance, in that educational goals and desired outcomes have seldom been reflected in pay for teachers and budgets for schools. Traditional teacher salary schedules provide higher pay for experience and postgraduate degrees, neither of which appears to be systematically linked with student achievement. Skill and knowledge-based pay shows greater promise for making teachers more effective in the classroom but remains to be tested. School-based accountability and incentive systems are increasingly popular and seem to contribute to desired student outcomes. To be fully effective, however, they require adequate funding for schools and attention to capacity building.

- Empowering schools or parents to make decisions about public funds (via enhanced site-based management, charter schools or contract schools, or vouchers, for example) has been justified as a strategy for improving student achievement in a cost-efficient way based on a variety of different arguments: some contend that local control will enhance innovation at the school level; some believe that schools with a strong sense of community perform better; and some believe that the introduction of competition and the possibility of losing students (and their associated funding) will encourage schools to be more productive than under the current monopoly situation. Although positive effects for children using vouchers have been reported from several sites where vouchers have been tried, the small scale of current programs leaves many important questions unanswered.

Although positive effects for children using vouchers have been reported from several sites where vouchers have been tried, the small scale of current programs leaves many important questions unanswered.

Achieving Goal 2: Reducing the Nexus Between Student Achievement and Family Background Characteristics

- As money is made to matter more in education, funding disparities will become increasingly worrisome, because their effects on achievement will be magnified to the detriment of children in underfunded schools, many of whom are likely to be from disadvantaged backgrounds. The new focus on funding adequacy has the potential to help disadvantaged students, but it will do so only to the extent that school funding formulas are appropriately adjusted for the additional costs of educating youngsters from disadvantaged backgrounds.

- Achieving goal 2 will also require attention to increasing both the capacity of children to learn and of schools to teach. Children raised in economically and socially impoverished environments or suffering from physical disabilities often come to school less ready to learn than their more advantaged counterparts. Schools must deal with these problems, even though they alone will not be able to solve them. A strong consensus has emerged among policymakers, practitioners, and researchers about the importance of increasing investments in the capacity of at-risk children to learn, by focusing on the school-readiness of very young children and by linking education to other social services, so that the broad range of educational, social, and physical needs that affect learning are addressed. Programs providing early childhood interventions and school-community linkages give evidence of both promise and problems, suggesting that there is still much to learn about making these investments effectively.

- That more investment is needed in the capacity of schools to educate concentrations of disadvantaged students would seem to be obvious given the dismal academic performance of many of these students, but as yet we have only incomplete answers to the question of which types of investments are likely to be the most productive and how to structure them to make them effective. The quality of teachers is likely to be a key compo-

nent; reducing class size might help under certain conditions; whole-school restructuring may have significant potential; and the dilapidated state of school buildings in many older urban areas suggests that reform of facilities financing must also be attended to. Again, the effectiveness of any individual policy change may depend on how it is linked to an interconnected set of strategies for improving school performance, and some critics question whether these most troubled of U.S. schools can be reformed through strategic investments and related strategies, or whether they require much more fundamental structural change, such as might be brought about by a voucher program.

- Most federal and some state aid flows to schools via categorical programs tied to the special needs of certain groups of disadvantaged students. Title I compensatory education grants and special education funding are the chief examples. Questions have been raised about the extent to which the incentives deliberately or inadvertently created by categorical programs serve educationally desirable purposes and whether and to what extent it continues to be appropriate to treat children with special needs separately in an educational system increasingly oriented toward fostering higher levels of learning for all students. Our findings suggest that previously defined sharp distinctions between students with special educational needs and other students have compromised educational effectiveness and that current efforts to move toward more integrated school programs should be facilitated by the finance system.

Most federal and some state aid flows to schools via categorical programs tied to the special needs of certain groups of disadvantaged students.

- Arguments for dramatic changes in school governance (by empowering schools or parents to make decisions about public funds) may be more compelling in urban areas with large numbers of disadvantaged students than in the educational system in general for a number of reasons. The size of many urban districts and the continuing fact of racial and economic segregation offer many urban residents much less choice over where and how to educate their children than suburban residents have. Moreover, urban residents have arguably benefited least from prior school reforms. Some economic models suggest that, among choice options, charter schools and vouchers, rather than interdistrict and intradistrict choice programs, are the approaches most worthy of further exploration as vehicles for improving poor-performing schools. At present, however, little is known about the effects of either. Extensive evaluation is needed of the many charter efforts currently under way. Vouchers, both publicly and privately funded, are being tried in a number of cities, but the existing small-scale efforts are unlikely to provide adequate information to assuage the concerns of those who question the need for so dramatic a break with traditional school finance policies.

Achieving Goal 3: Raising Revenue Fairly and Efficiently

- Shifting away from local revenue raising to greater reliance on state revenues and/or increasing significantly the federal role in revenue provision for elementary and secondary education would foster the goal of raising revenues fairly. Both, however, have to be considered in light of trade-offs and complementarities with the other two goals of a good financing system and with attention to maintaining some local control over managerial decisions.

- A larger federal role in providing education revenues could be justified either on the grounds that is fair and appropriate for the federal government to take responsibility for disproportionate needs of students who are poor, who have disabilities, or are otherwise educationally disadvantaged, or on the grounds of ensuring that all states can provide adequate education funding. Fully funding federal compensatory education programs would be consistent with past federal policy and is likely to be the more politically viable of the two approaches. The alternative of a new federal foundation aid program based on an adequacy justification would entail a significant change in federal policy and would raise many of the same analytical, conceptual, and political issues that arise in

the formulation of adequacy programs at the state level.

Finally, the report draws attention to the nation's need for better and more focused education research to help strengthen schools and bring about substantial improvements in student learning. Acknowledging the especially challenging conditions facing many big-city educators, the committee proposes three new substantial research initiatives in urban areas (without specifying the priority among them): (1) an experiment on capacity-building that would tackle the challenges of developing and retaining well-prepared teachers; (2) systematic experimentation with incentives designed to motivate higher performance by teachers and schools; and (3) a large and ambitious school voucher experiment, including the participation of private schools. Meeting the nation's education goals will depend in part on continuously and systematically seeking better knowledge about how to improve educational outcomes, through new research initiatives such as these along with more extensive evaluation of the many reform efforts already under way.