NCES
National Center for Education Statistics

U.S. Department of Education
Institute of Education Sciences
NCES 2004-325

Developments in School Finance: 2003

Fiscal Proceedings From the Annual State Data Conference of July 2003

NCES
National Center for
Education Statistics

U.S. Department of Education
Institute of Education Sciences
NCES 2004–325

Developments in School Finance: 2003

Fiscal Proceedings From the Annual State Data Conference of July 2003

August 2004

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

U.S. Department of Education
Rod Paige
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Robert Lerner
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
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006–5651

August 2004

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

Suggested Citation

Fowler, W.J., Jr., ed., (2004). *Developments in School Finance, 2003: Fiscal Proceedings From the Annual State Data Conference of July 2003,* (NCES 2004–325). U.S. Department of Education, National Center for Education Statistics, Washington, DC: Government Printing Office.

For ordering information on this report, write:

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

call toll free 1–877–4ED–PUBS, or order online at www.edpubs.org

Content Contact

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

Foreword

Jeffrey A. Owings

Associate Commissioner
NCES Elementary/Secondary and Libraries Studies Division

At the 2003 National Center for Education Statistics (NCES) Summer Data Conference, scholars in the field of education finance addressed the theme, "Data Changing Our World." Discussions and presentations dealt with topics such as the effects of salary and working conditions on teacher turnover, determining the cost of improving student performance, and measuring school efficiency.

Developments in School Finance: 2003 contains papers presented at the 2003 annual NCES Summer Data Conference. The presenters are experts in their respective fields, each of whom has a unique perspective or who has conducted 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 volume is the eighth education finance publication produced from papers presented at the NCES Summer Data Conferences. The papers included in this volume 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 suggestions and comments of the reviewers at the National Center for Education Statistics (NCES): Jeffrey Owings, Associate Commissioner for the Elementary/Secondary and Libraries Studies Division, who provided overall direction; and Tai Phan, who provided technical review of the entire publication. At the Education Statistics Services Institute (ESSI), Leslie Scott provided technical review of the publication, and Tom Nachazel proofread and coordinated production of the publication, with assistance from other members of the ESSI editorial team. Also at ESSI, Heather Block and Jennie Jbara performed the desktop publishing.

Contents

Introduction

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

The papers included in this volume of fiscal proceedings were presented by education finance experts at the July 2003 NCES Summer Data Conference. The presenters were invited by the editor to contribute their papers to this volume because, in his opinion, their work in elementary-secondary public school education finance is among the leading work in the field. The following paragraphs present an overview of each of the papers in this volume, in the order in which they appear. For each paper, the title (in bold) and list of authors and their affiliations introduce the paper summary.

The Revolving Door: Factors Affecting Teacher Turnover. In this paper, Eric A. Hanushek of the Hoover Institution at Stanford University, the late John F. Kain of the University of Texas at Dallas, and Steven G. Rivkin of Amherst College use Texas teacher data to conclude that Texas public school teachers' working conditions matter more to them than salary. The authors state that although experienced teachers, on average, are more effective at raising student performance, they typically either leave teaching or flee from urban to suburban schools. What has not been well understood, the authors assert, is whether experienced teachers leave schools with high concentrations of disadvantaged and low-achieving students for reasons of compensation, or because of their working conditions.

Hanushek, Kain, and Rivkin state the reason this issue has not been resolved has been the difficulty of separating the effects of teachers' salary from the effects of their working conditions and preferences. This requires databases with detailed information on enough teachers and students to statistically distinguish what influences teachers' decisions. Using data from the state of Texas for elementary schools from 1993 to 1996, the researchers were able to construct such a database.

The authors report that teachers leave teaching or transfer from one school to another in response to the characteristics of their students more than better salaries in other schools. They posit that this is why disadvantaged, low-achieving students are in schools with relatively inexperienced teachers. Since salary does not seem to be the primary motivation for exiting, the authors suggest that improving the working conditions in these inner-city schools, as well as increasing the salaries of "quality" teachers, should be considered by policymakers.

Financing Urban Schools: Emerging Challenges for Research, Policy, and Practice. In this paper, Christopher Roellke of Vassar College and Jennifer King Rice of the University of Maryland review contemporary research on the financing of urban school systems. The dilemma for the nation's large urban schools, Roellke and Rice posit, is that they are particularly vulnerable to funding reductions as states and localities respond to lower revenues and deficits, at the same time that they are the focus of increased standards and accountability. In part, the increased demands are because policymakers wish to close the achievement gap between low-income, minority students and more fortunate, better achieving students. The authors, in the first volume of a new book series on education fiscal policy and practice, asked a group of school finance experts to address the critical challenges in financing urban education, and they synthesize the key themes that arose.

Roellke and Rice report that the solutions to the funding problems of urban schools are not easily addressed by research. Currently, urban schools have many service delivery options available to them. Examples include implementing class-size reduction, alternative scheduling, summer enrichment, early intervention programs, and a wide array of whole-school reform models. To date, education finance researchers have little to offer on the cost or the effectiveness of alternative practices to assist urban policymakers in choosing between service delivery options. In addition, it is apparent that no one solution can improve student outcomes in all schools. One contributor to the volume, Jennifer Imazeki, states that additional compensation alone appears to be insufficient to attract and retain high-quality teachers in urban schools. And the vast majority of public schools, according to contributors Schwartz, Amor, and Fruchter, receive private support, which for some urban schools represents over half of their children services funding. Thus, Roellke and Rice argue, the ability of some urban schools to implement certain reforms is dependent upon this episodic nontraditional funding.

Roellke and Rice also report that urban school districts do not have the resources to close the achievement gap between their low-income, minority students and more fortunate, better achieving students in other school districts. This inability has resulted in legal challenges in state courts, focused on allegations that the state education finance formula does not allow poor school districts to provide an adequate education. The authors state that some state courts, such as in New York, have required a costing-out study to set a dollar figure that would ensure an adequate education.

Financing Education So That No Child Is Left Behind: Determining the Costs of Improving Student Performance. In this paper, Andrew Reschovsky of the University of Wisconsin-Madison and Jennifer Imazeki of San Diego State University estimate the cost of achieving a specified improvement in student performance. To do this, they use the characteristics of schools and students in Texas that may cause some schools to spend more than others to achieve a given student performance standard. The authors state that these costs are due primarily to factors over which local school officials have little control, such as high concentrations of low-income students or ESL students who may require, for example, smaller classes or specialized programs. In addition, the authors state, some schools, because of their location and student composition, may have to offer higher compensation to attract and retain staff.

Reschovsky and Imazeki suggest that substantial cost differences among school districts will render those with above-average costs unable to bring their students up to the new standards, unless these school districts receive additional aid. The authors report palpable cost differences in Texas. They then devise a cost-adjusted foundation formula for the state to send more funds to school districts with higher costs.

Reschovsky and Imazeki caution that their estimated cost functions should not be interpreted to mean that if a school district with high costs is provided with sufficient additional funds it could meet state-imposed performance standards in a single year. It may take more time than anticipated for a school district to reach any specified state standard, particularly if the school district is substantially below the desired standard. In addition, the authors state that a one-time increase in state aid would not be as effective as a gradual phase-in.

Distinguishing Good Schools From Bad in Principle and Practice: A Comparison of Four Methods. In this paper, Ross Rubenstein of the Maxwell School of Citizenship and Public Affairs at Syracuse University and Leanna Stiefel, Amy Ellen Schwartz, and Hella Bel Hadj Amor of the Robert F. Wagner Graduate School of Public Service at New York University compare four quantitative techniques to measure school performance. The distinctive aspect of this paper is that the authors estimate efficiency scores, particularly the identification of "good" and "bad" schools, using each of four methods based on the same data. They use New York City and Ohio school data that includes student characteristics, test scores, and school resources. They then explore how and why the methods and results differ. The four efficiency techniques explored in the paper are adjusted performance measures (APMs); data envelopment analysis (DEA); education production functions (EPFs); and cost functions.

The authors assert that one of the most difficult challenges is comparing schools educating diverse students, particularly those without the necessary resources. Variations in student performance are highly correlated with students' socioeconomic backgrounds. Urban schools serving primarily minority and low-income students may appear to be low performing principally as a result of factors outside their control. What the authors wish to do is to find schools that make the most effective use of their limited resources, and explore how they make use of those resources.

Even using the same data and specifications, the authors demonstrate that different analytic methods may also produce different results. The different analytic methods, however, do produce lists of "good" and "bad" schools that are similar. And the use of more than one analytic method improves the accuracy of the identification of schools as either "good" or "bad." The authors state that

> . . . simplistic measures of school performance, which do not account for the complex environment of schooling, risk identifying the wrong schools as either exemplars or in need of interventions. This problem is particularly critical when the performance measures are used to distribute rewards and sanctions.

Court-Mandated Change: An Evaluation of the Efficacy of State Adequacy and Equity Indicators. In this paper, Jennifer Park and Ronald A. Skinner of *Education Week* explore the validity of their equity and adequacy grades for states' school finance systems. They examine four states—New Hampshire, New Jersey, Vermont, and Wyoming—that have recently changed their school finance system as a result of losing a court challenge to the state education funding system. The authors compare the main equity and adequacy indicators before and after court-mandated changes were implemented. What the authors sought to do was to verify that these indicators were accurately portraying the changes that were occurring in these state education finance systems.

In order to test the validity of the equity and adequacy measures in the four states, information regarding the school finance and litigation history of these four states was collected by the authors from court decisions, legislative changes to the state education funding system, and published studies.

Park and Skinner report that the indictors for two of the four states they examined, New Hampshire and New Jersey, matched very well with actual change. Vermont and Wyoming's changes were less clear, perhaps, they state, because the reforms in these states were implemented over several years and this analysis looked for changes in indicators in the single year where the most changes occurred. Some indicators were more accurate than others, and the authors suggest using a weighting system that more heavily weights the more accurate indicators. Because of the time lag in the availability of federal school finance data, the authors emphasize that current contextual information is important to consider when interpreting these equity and adequacy indicators.

School Finance Reform and School Quality: Lessons From Vermont. In this paper, Thomas Downes of Tufts University examines the changes in Vermont's distributions of education spending resulting from the 1997 enactment of Act 60. Specifically, Downes examines whether the resulting changes in the distributions of spending have generated greater equality in measured student performance. Act 60, the "Equal Educational Opportunity Act," may be the most radical reform of a state's system of public school financing since the post-*Serrano* and post-Proposition 13 changes in California in the late 1970s, according to Downes. Prior to Act 60, Vermont used a traditional foundation formula to give towns state education aid. Act 60 created a combined foundation and power equalization plan that included a statewide property tax. The changes, which Downes states were designed to shift some of the burden away from state residents to corporations and nonresident owners, were phased in over several years.

Examining local education spending inequality before and after enactment of Act 60, Downes finds that inequality generally declined. Examining the relationship of spending and town wealth, he finds that wealthier towns did spend more prior to Act 60. However, Downes asserts that care must be taken not to make too much of the declines in inequality, as they are small and not consistent. The small declines in the disparities in student performance would not, he states, justify a major policy change like Act 60.

Downes concludes that Act 60 was a dramatic change in Vermont's education funding, and that his analyses demonstrate a reduced range in education spending resulting from weakening the link between spending and property wealth. In addition, Downes tentatively concludes there is some evidence that student performance has become more equal since enactment of Act 60, but the improvements have been small.

Shopping for Evidence Against School Accountability. In this paper, Margaret E. Raymond and Eric A. Hanushek of the Hoover Institution at Stanford University explore whether or not accountability is associated with more gains in learning by students. The authority behind accountability has spread from states to also include the No Child Left Behind Act of 2001 (NCLB), which mandates reporting and accountability through testing. Opponents, Raymond and Hanushek assert, aggressively search for evidence that testing and accountability are actually harmful to students. The existing evidence on state accountability systems, Raymond and Hanushek report, indicates that their use leads to better student achievement.

Raymond and Hanushek use "The Nation's Report Card" (National Assessment of Educational Progress) scores to demonstrate that states with their own "report cards," which serve as a public disclosure, without sanctions and rewards, have gains of up to 1.2 percent. Those states that disclose publicly and also have sanctions and rewards (which the authors call "accountability" systems) have a 1.6 percent increase in mathematics performance. However, the introduction of an accountability system also has unintended consequences, they state, such as cheating. These findings differ, however, from those of previous studies by other researchers.

The conclusion reached by Raymond and Hanushek is that the media is unaware of or indifferent to quality differences in the competing evidence on accountability program performance. Since the media, many policymakers, and decisionmakers in education agencies serve a "gatekeeper" function for disseminating information to the general public, evidence quality is of great importance, Raymond and Hanushek assert. When millions of dollars are involved, as they are in accountability systems, evidence must meet the highest scientific standards, the authors declare, reliably controlling for rival alternative explanatory factors.

Raymond and Hanushek conclude that no one yet understands how best to design accountability systems that can be directly linked to incentive systems.

The Revolving Door: Factors Affecting Teacher Turnover

Eric A. Hanushek
Stanford University

John F. Kain
University of Texas at Dallas

Steven G. Rivkin
Amherst College

About the Authors

Eric A. Hanushek is the Paul and Jean Hanna Senior Fellow at the Hoover Institution of Stanford University, chair of the executive committee for the Texas Schools Project, and a research fellow of the National Bureau of Economic Research. He has written extensively about the economics and finance of schools. He can be contacted at hanushek@stanford.edu.

Steven G. Rivkin is an associate professor of economics at Amherst College, Associate Director of Research for the Texas Schools Project, and a research fellow of the National Bureau of Research. He has written extensively about the economics and sociology of education. He can be contacted at sgrivkin@amherst.edu.

The late John F. Kain was a professor of economics and political economy at the University of Texas at Dallas. After a career studying urban housing and location as professor of economics at Harvard University, he moved to Texas and founded the Texas Schools Project.

The Revolving Door: Factors Affecting Teacher Turnover

Eric A. Hanushek
Stanford University

John F. Kain
University of Texas at Dallas

Steven G. Rivkin
Amherst College

Acknowledgments

This essay is adapted from articles in the winter 2004 issue of *Education Next*[1] and the spring 2004 issue of the *Journal of Human Resources.*[2]

Introduction

Experienced teachers are, on average, more effective at raising student performance than those in their early years of teaching.[3] This gives rise to the concern that too many teachers leave the profession after less than a full career and that too many leave troubled inner-city schools for suburban ones. Until now, the roots of these problems have not been well understood. In particular, it is not known whether teachers leave schools with high concentrations of disadvantaged and low-achieving student populations for financial reasons or because of the working conditions associated with serving these students. Nor are there good estimates of what kinds of salary increases would need to be offered to slow the turnover among teachers.

Significant policy decisions rest on understanding the market for teachers and the responsiveness of teachers to varying conditions of employment. This article summarizes our recent analysis of teacher decisions. The underlying technical article (Hanushek, Kain, and Rivkin [2004b]) provides detailed statistical estimates of teacher behavior. Here we distill the main findings with an eye toward the policy implications of that work.

The chief obstacle to resolving these issues has been the difficulty of separating the effects of teachers' salary levels from their working conditions and preferences. The outstanding suburban school that retains most of its teachers is likely to be attractive on a number of levels: the pay is good, students are high performing,

---

[1] Hanushek, Kain, and Rivkin (2004a).

[2] Hanushek, Kain, and Rivkin (2004b). This article is the technical version of this essay.

[3] Our work on teacher quality differences finds that teachers in their initial years of experience perform significantly poorer than later in their career. This effect appears to be primarily a "learning effect" where teachers improve through on-the-job experience, as opposed to a compositional effect arising from the fact that many, possibly less skilled, early career teachers tend to exit teaching altogether. See Rivkin, Hanushek, and Kain (2001) and Hanushek and Rivkin (2004).

and parents are supportive. Since all three factors help in attracting and retaining teachers, it becomes difficult to calculate the degree to which each factor separately affects a teacher's decision to return to that school the following year. Conversely, the school that has disadvantaged and low-performing students may suffer high rates of teacher turnover, but sorting out the causes of turnover is difficult. Doing so requires detailed information for enough teachers and students to allow analysts to distinguish statistically among the various factors that affect teachers' decisions.

Fortunately, important parts of the necessary information are now available for elementary schools in the state of Texas for the years 1993 through 1996. Working in cooperation with the Texas Education Agency, the University of Texas at Dallas's Texas Schools Project has combined various data sets to create a database of key characteristics of both teachers and students during this period in all Texas public schools. This information includes the race, ethnicity, and gender of both students and teachers; students' eligibility for a subsidized lunch; and students' performance on the Texas Assessment of Academic Skills (TAAS), a criterion-reference test administered each spring to students in grades 3 through 8. The database also contains annual information about the teachers: their years of experience, their education and salary levels, the grades and subjects they teach, and the size of their classes.

Our analysis of these data reveals that teachers transfer from one school to another—or exit the Texas public school system altogether—more as a reaction to the characteristics of their students than as a response to better salaries in other schools. This tends to leave disadvantaged, low-achieving students with relatively inexperienced teachers. Because teachers appear so unresponsive to salary levels, it would take enormous across-the-board increases to stem these flows. Indeed, the results suggest that policymakers ought to consider only selective pay increases, preferably keyed to quality, for work in inner-city schools, together with efforts to improve the working conditions in these schools.

Reasons for Leaving

Teachers decide whether to remain at a school for a multiplicity of reasons, which can be divided into four main categories: (1) characteristics of the job, including salary and working conditions; (2) alternative job opportunities; (3) teachers' own job and family preferences; and (4) school districts' personnel policies. Although we were not able to look at the ways in which all of these factors affect teachers' decisions with respect to their employment situation, we were able to examine directly the impact of salary and certain working conditions. We were also able to draw some reasonable inferences about how family considerations and alternative job opportunities influence teachers' decisions by examining how their choices differ by gender and experience.

Teachers transfer from one school to another more as a reaction to the characteristics of their students than as a response to better salaries.

Admittedly, "working conditions" is a broad concept that can cover everything from class size to discipline problems to student achievement levels. Though we do not have observational data on every aspect of teachers' working conditions, we do know certain characteristics of their students that many believe affect the teaching conditions at a school: the percentage of low-income students at the school (as estimated by the percentage eligible for a subsidized lunch), the shares of students who are African American or Hispanic, average student test scores, and class sizes. Whether these characteristics directly affect teachers' decisionmaking or indicate other less tangible factors (such as the disciplinary climate or bureaucratic environment at the school) cannot be determined.

When looking at the impact of working conditions on retention rates, one needs to take into account other factors that may affect teachers' employment choices. Some teachers possess skills that are considered more valuable in the private sector employment marketplace. For instance, mathematics and science teachers may find more demand for their services in the private sector than an English teacher would. However, our study focuses on elementary school teachers, who tend to have similar educational backgrounds and similar opportunities outside the education system. As a result, differences in private sector alternative employment opportunities among teachers of different subjects should not be very important for this analysis.

A more important consideration is that many teachers may wish to remain at a particular location for other than job-related reasons, perhaps out of a desire to live near their hometown or near their spouse's workplace. Consequently, the availability of jobs in the locality may be an important determinant of the probability of exiting a school, and we control for any systematic differences across regions within Texas.

Retention rates can also be affected by the number of years a teacher has spent in a particular location. The more years working in a particular district, the more costly it becomes to leave, simply because pay, responsibilities, and job opportunities are often tied directly to experience within the same school district. The financial attractiveness of moving elsewhere also attenuates with the passage of time. Because many districts credit a transferring teacher with only a limited number of years of experience, teachers may have to take a salary cut if they switch school districts. In general, switching careers grows costlier with age and experience. One must give up the higher salary that comes with experience within a particular field, and the time to accumulate gains from any change in job or career grows shorter as one ages. For this reason, our analysis takes into account the number of years teachers have held their jobs by comparing only teachers with similar levels of experience.

Each year, approximately one-fifth of all teachers nationwide decide to leave the school at which they are teaching.

Other relevant differences among teachers may arise from their family circumstances, such as the job opportunities of a spouse or a desire to stay home with young children or to enjoy the benefits of home ownership. For example, many female teachers who leave teaching do so in order to leave the labor market altogether, often for family reasons. We unfortunately lack information on family structure, sources of income other than salary, the location or type of housing, and whether and where a spouse works. However, we are able to look separately at teachers grouped by gender, giving us an opportunity to assess the extent to which female and male teachers are influenced by different school-related factors.

Ethnicity may also affect decisionmaking. Teachers may prefer to teach in schools where they share the ethnic characteristics of the students, or they may find it easier to obtain a position if administrators prefer instructors who have certain ethnic characteristics. To ascertain whether ethnic background affects teachers' decisionmaking, we also look separately at White, African American, and Hispanic teachers.

One limitation of our study is that we do not have direct information on school districts' hiring and retention practices. Districts have options when hiring, and the willingness of a teacher to leave a position will depend on the availability of an attractive position elsewhere. Although few teachers are involuntarily separated from their jobs, we do not know whether a job change is determined primarily by a teacher's decision or by that of the employer, and the circumstances undoubtedly affect both opportunities and the range of choices a teacher will consider. Our lack of information about employer-initiated moves may lead to an underestimate of the improvements in pay and working conditions achieved by teachers who move voluntarily, but the size of this underestimate is probably not very large.

Movement Between and Within Districts

Each year, approximately one-fifth of all teachers nationwide decide to leave the school at which they are teaching. The pattern in Texas is roughly the same as in the nation as a whole. On average, in each year between 1993 and 1996, more than 18 percent of Texas teachers decided not to remain at the school at which they were teaching. More than 6 percent changed schools within their districts, another 5 percent switched from one district to another, and 7 percent left Texas public schools altogether.

Let's look first at the changes in salary typically experienced by teachers moving to a new district. Instead of relying on salary data reported for each individual teacher, we calculate district average salaries for teachers in each of their first 10 years of experience during the period from 1993 to 1996. These averages are based on regular pay for teachers without advanced degrees and exclude extra pay for coaching or other activities. (The latter is not an important part of compensation: Over 85 percent of teachers receive no extra pay, and the median extra pay for those who do receive it is about $1,000 per year.) We use these averages to characterize the salary schedule of each district and then estimate the potential salary change resulting from a move, given the experience level of each teacher. For example, the salary change for a teacher who switches districts after 4 years of teaching is assumed to equal the average salary of fifth-year teachers in the new district minus the salary for that level of experience in the old district.

On average, teachers who move between districts after no more than 2 years at a school improve their salaries, though just barely. Male teachers gain 1.2 percent in salary, while women gain 0.7 percent. Even these small gains begin to disappear for teachers with more experience. Overall, the average annual salary gain among all teachers with less than 10 years' experience is 0.4 percent of annual salary, or roughly $100. Women with 3 to 9 years of experience who decide to change districts actually take, on average, a small pay cut. In short, most teachers moving between districts do not receive substantially better pay in their new jobs.

Teachers moving between districts have the opportunity to teach higher achieving, higher income, nonminority students.

The picture for working conditions is quite different. There is strong evidence that teachers moving between districts have the opportunity to teach higher achieving, higher income, nonminority students. The findings for achievement are the clearest and most consistent. The average job switcher moving from one district to another moved to a district whose average achievement was 0.07 standard deviations higher on the Texas Assessment of Academic Skills than that of the district the teacher left. (The difference is 3 percentile points on a 100-point scale.) The shares of the district's students who were African American, Hispanic, or low income also declined significantly for movers. On average, the districts to which teachers moved had 2 percentage points fewer African American students, 4.4 percentage points fewer Hispanic students, and more than 6 percentage points fewer low-income students (of any ethnicity) than the districts they had left.

These patterns were even more pronounced for teachers who moved from urban to suburban districts. The salaries of such teachers actually declined by 0.7 percent, on average, as a result of their moves. Meanwhile, the average achievement in the new districts increased by 0.35 standard deviations (14 percentile points), and the shares of African American and Hispanic students decreased by 14 and 20 percentage points, respectively. Teachers who moved between different suburban districts experienced similar, albeit smaller, changes in student characteristics. Student achievement in their new districts was one-tenth of a standard deviation higher, while the percentages of African American, Hispanic, and economically disadvantaged students all declined.

We can gain further insight into the factors associated with teacher mobility by examining the pre- and post-move school characteristics for teachers moving to a new school within the same district. These results confirm that teachers who move between schools within urban districts typically arrive at a school with higher average student achievement (0.11 standard deviations) and a smaller percentage of minority and low-income students. In other words, those who choose to change schools within districts appear to follow the same attributes, seeking out schools with fewer academically and economically disadvantaged students. These patterns are also consistent with the notion that new teachers are often placed in the most difficult teaching situations and that senior teachers can often choose more comfortable positions within the system.

Important differences emerge, however, when we separate teachers by their own ethnic background. African American teachers tend to move to schools with *higher* percentages of African American enrollment than their previous schools, regardless of whether they change districts or simply move to a new school in the same district. However, the average change in the percentage of Hispanic students for teachers of Hispanic descent is not much different from the changes experienced by teachers as a whole. The typical gap in average test scores between their current and former school is also much smaller for African American and Hispanic teachers who have switched schools.

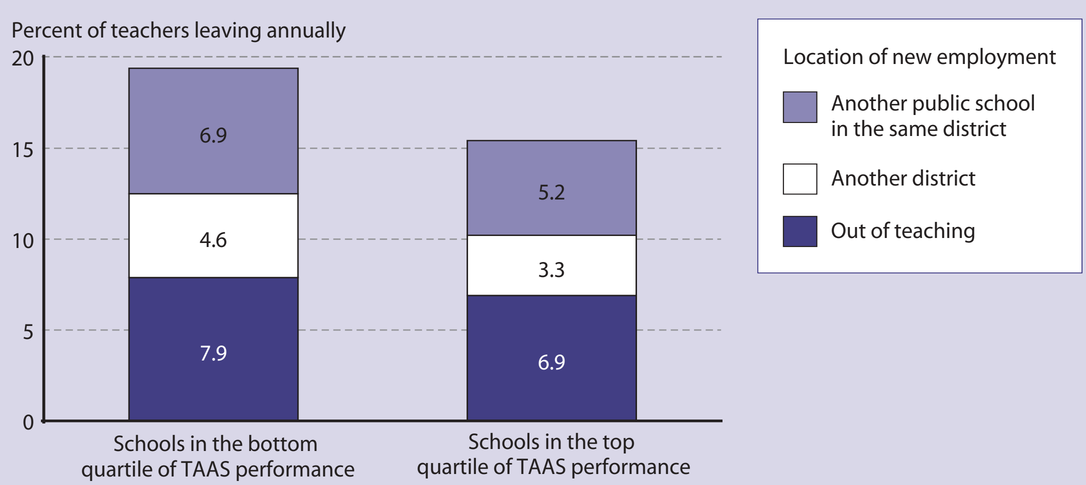
It is not clear whether these ethnic differences are the result of teachers' preferences or of the job opportunities available to them. It could be that African American teachers prefer to work at a school near where they live. If so, then residential segregation by race may lead to their selection of schools with more African American students. Or teachers may simply prefer to teach students of a similar ethnic background. Alternatively, job opportunities for African American teachers may be more extensive in schools with higher proportions of African American students.

All this movement of teachers among schools obviously affects the composition of the teaching force at particular schools. Since exiting rates are smaller at schools with more advantaged students, these schools also enjoy more experienced teachers. The pattern is particularly striking when schools are grouped according to their average level of student achievement. As figure 1 shows, almost 20 percent of teachers in schools in the bottom quartile of student achievement leave their schools each year, while only 15 percent of teachers in schools in the top quartile leave their schools each year. The driving force of this relationship is not simply teachers' leaving urban districts for suburban ones; more of the difference in leaving rates between these types of schools is caused by teachers moving to new schools within their original district, and there are nontrival differences in the rates of leaving teaching entirely. Since teachers with fewer than 2 years of experience tend to be less effective than more experienced teachers, existing mobility patterns in Texas are likely to adversely affect the achievement of disadvantaged students.

Salaries and Student Demographics

The analysis to this point has not disentangled the effects of salaries from the effects of the working conditions associated with students of varying achievement and family backgrounds. To identify more precisely the independent effects of the multiple factors affecting teachers' choices, we use regression analysis to estimate the separate effects of salary differences and school characteristics on the probability

Figure 1. Yearly percentage loss of teachers, by school's performance on the Texas Assessment of Academic Skills (TAAS) and location of teacher's new position



SOURCE: Authors.

that a teacher will leave a school district in a given year, holding constant a variety of other factors, including class size and the type of community (urban, suburban, or rural) in which the district is located. We also compare the impact of salaries and school characteristics on the probability of switching to another district with their impact on the probability of leaving teaching altogether.

The results of this analysis confirm that teachers are more likely to leave districts with low average achievement scores. Ethnic composition of the student body is also an important determinant both of the probability of leaving the public schools entirely and of switching from one school district to another. White teachers, regardless of their teaching experience, will tend to move to schools with fewer African American and Hispanic students. Less experienced White teachers are also more likely to leave the public schools altogether if they come from schools with higher concentrations of African American and Hispanic students. African American and Hispanic teachers, however, do not show the aversion to concentrations of minority students.

Although the ethnic composition of the school is the most important factor affecting teachers' decisions to change jobs, financial considerations are also relevant.

The differential effect of the ethnic composition of the student body for White and African American teachers could reflect personnel policies that prefer minority teachers in schools with higher concentrations of minority students. But teachers' own preferences may be even more important, as suggested by the fact that the decision to leave the Texas public schools altogether—a decision much more closely related to the individual teacher's preferences than to the district—is influenced in the same way by the schools' ethnic composition.

Although the ethnic composition of the school is the most important factor affecting teachers' decisions to change jobs, financial considerations are also relevant, especially when it comes to a decision by a male teacher to move from one district to another. For male teachers with fewer than 3 years of experience, the estimated change in the probability of switching districts for a 10 percent increase in salary is 2.6 percentage points; for men with 3 to 5 years of experience, the estimated change for a salary increase of the same magnitude is 3.4 percentage points; for still more experienced male teachers, financial effects trail off, down to essentially zero for those with more than 20 years of experience.

The corresponding numbers for less experienced women teachers are less than half those for men. Moreover, salary differences have no observable effects on women with 6 or more years of experience. The unresponsiveness of female teachers to salary increases is important in the subsequent policy discussion, since female teachers represent the vast majority of elementary school teachers.

Policy Implications

The results presented above confirm the difficulty that schools serving academically disadvantaged students have in retaining teachers, particularly those early in their careers. Teaching lower achieving students is a strong factor in decisions to leave Texas public schools, and the magnitude of the effect holds across the full range of teachers' experience levels. There is also strong evidence that a higher rate of minority enrollment increases the probability that White teachers will leave a school. By contrast, increases in the shares of African American and Hispanic students reduce the probability that African American and Hispanic teachers will leave.

Given these findings, a key question is how to reduce the flows out of low-achieving, high-minority schools and out of the teaching profession altogether. One oft-proposed solution is to provide teachers with "combat pay"—salary increments designed to encourage them to remain at a tough school. But how large would the increase need to be in order to neutralize the effects of difficult working conditions? Let's consider this closely.

The situation is complicated by the fact that most elementary school teachers in Texas are White females (only 20 percent are African American or Hispanic, while only 14 percent are male). As noted earlier, female teachers are less responsive to increases in salary, meaning that the bonus required to keep them at a school will be larger than for males. In addition, White teachers are the most likely to exit low-achieving, high-minority schools, meaning that it will take even larger increases to retain them. If the teaching corps looked much different—say, if the teachers in urban elementary schools were mostly African American and Hispanic males—the costs of the "combat pay" solution would be lower.

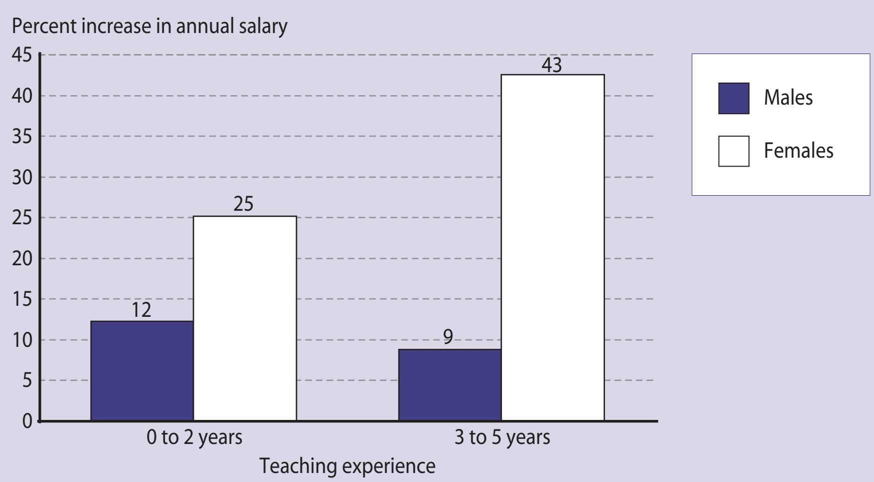
Based on our findings of what causes teachers to leave their schools, we calculated the salary increases that would be necessary to offset the effects of difficult working conditions in large urban versus suburban schools.[4] These calculations, performed separately for White male and female teachers in their early careers, are shown in figure 2. The findings suggest that truly large boosts in salary would be needed, particularly among women. Female teachers in large urban school districts would require a 25 percent initial increase in compensation, rising to over 40 percent when they reach 3 to 5 years of experience. Moreover, this is only in the "typical" urban school. For the neediest or most troubled schools in urban areas, even the differentials calculated in figure 2 would probably not be sufficient to stem the high levels of turnover in such schools.

Across-the-board salary increases of 25 to 40 percent for teachers in urban areas would be an enormously expensive reform, and, in addition, it would be difficult to target such a solution, since teachers typically negotiate salary schedules that apply to all the teachers in the district, not just to those in the most disadvantaged schools. Similarly, even if targeted to the most disadvantaged schools, any increases in salaries would almost certainly go to new and middle-career teachers alike, even though we know that salary differentials are nearly irrelevant for women teachers with 10 or more years of experience.

At this time, we do not fully understand the working conditions that are most important, but we might speculate that at least a component involves school

[4] These calculations, described in Hanushek, Kain, and Rivkin (2004b), rely on the estimated effects of salary and student characteristics on the probability of leaving a school. From the differences in student characteristics for the average urban and average suburban school, we calculate the increased probability that a teacher (of a given gender and experience category) will leave urban schools. Then, based on the impact of salaries on exit rates, we calculate the salary premium needed to neutralize the effect of these adverse student characteristics.

Figure 2. Increase in salary of nonminority urban teachers necessary to equalize turnover between urban and suburban schools, by experience and gender of teacher



NOTE: Estimates based on the differences in average achievement and in the shares of African American and Hispanic students between large urban and suburban districts.

SOURCE: Authors.

characteristics that are simply associated with the student characteristics we have identified. To the extent that other characteristics of schools where disadvantaged students are found—such as safety and disciplinary problems, more bureaucratic rules, poor leadership, greater student turnover, or a greater commuting distance—are important elements, improving these working conditions could mitigate the turnover problem we have identified. And these improvements might even have a direct effect on student performance.

Finally, it is important to note that this study focuses solely on how many teachers transition among schools and out of teaching. We have not examined the quality of the teachers who move from one district to another or leave teaching altogether. The actual cost of improving the quality of instruction depends crucially on whether good teachers, not just experienced teachers, are being retained. Salary policies that are guided just by the characteristics of the students in a school will retain both the good and the bad teachers.

We do know from our other work that differences in teacher quality are more significant than the differences arising from having inexperienced teachers.[5] Therefore, an approach with more appeal might be simply to accept the fact that there may be greater turnover in schools serving a larger disadvantaged population, but then to concentrate much more attention and resources on the quality dimension. While we do not have much experience with such policies, they seem like the most feasible way to deal with the problems of schools serving low-income and minority students.

[5] Rivkin, Hanushek, and Kain (2001) establish lower bounds on the variation in teacher quality that is found in elementary schools in Texas. That analysis suggests that perhaps 10 percent of the variation in quality arises from experience effects that will change as teachers pass their initial period of teaching.

References

Hanushek, E.A., Kain, J.F., and Rivkin, S.G. (2004a). The Revolving Door. *Education Next, 4*(1): 76–82.

Hanushek, E.A., Kain, J.F., and Rivkin, S.G. (2004b). Why Public Schools Lose Teachers. *Journal of Human Resources, 39*(2): 326–354.

Hanushek, E.A., and Rivkin, S.G. (2004). How to Improve the Supply of High Quality Teachers. In D. Ravitch, (Ed.), *Brookings Papers on Education Policy 2004* (pp. 7–44). Washington, DC: Brookings Institution Press.

Rivkin, S.G., Hanushek, E.A., and Kain, J.F. (2001). *Teachers, Schools, and Academic Achievement* (Rev. ed.) (NBER Working Paper No. w6691). Cambridge, MA: NBER.

Financing Urban Schools: Emerging Challenges for Research, Policy, and Practice

Christopher Roellke

Vassar College

Jennifer King Rice

University of Maryland

About the Authors

Christopher Roellke is associate professor and Chair of Education at Vassar College, where he teaches educational policy, the foundations of education, and urban education reform. His research interests are in the politics and economics of education, teacher education, and secondary school reform. Most recently, his work has focused on issues of teacher recruitment, induction, and retention in urban schools. He is co-editor (with Jennifer King Rice) of *Fiscal Policy in Urban Education* and has recently published articles in *The Journal of Education Finance*, *Education and Urban Society*, *The International Journal of Educational Research*, and *The Economics of Education Review.* He can be contacted at chroellke@vassar.edu.

Jennifer King Rice is associate professor in the Department of Education Policy and Leadership at the University of Maryland. Her research draws on the discipline of economics to explore education policy questions concerning the efficiency, equity, and adequacy of U.S. public education; her current work focuses on teachers as a critical resource in the education process. She is author of *Teacher Quality: Understanding the Effectiveness of Teacher Attributes*, and is co-editor (with Christopher Roellke) of *Fiscal Policy in Urban Education.* Her research has been published in numerous journals, including *Educational Evaluation and Policy Analysis*, *Economics of Education Review*, *Educational Policy, Education Administration Quarterly,* and *The Journal of Education Finance,* as well as multiple edited volumes. She can be contacted at jr162@umail.umd.edu.

Financing Urban Schools: Emerging Challenges for Research, Policy, and Practice

Christopher Roellke
Vassar College

Jennifer King Rice
University of Maryland

> *If school finance has seasons, then these are bitter cold days of winter for districts everywhere, even those in the sunniest of climates.*[1]

The uncertainty surrounding the current economic climate has an immediate and direct impact on education fiscal policy and practice. The nation's large urban education systems, serving high numbers of low-income, minority children, seem particularly vulnerable to programmatic cuts as states and localities respond to an environment of reduced revenues and huge budget deficits. Paradoxically, it is precisely these types of school systems that are the primary subjects of increased standards and accountability as policymakers and school officials seek to close the achievement gap between low-income, minority students and wealthier, predominantly White students.

Many urban districts across the country have invested heavily to meet new federal student achievement goals contained in the No Child Left Behind Act. These same districts now face the challenge of meeting these mandates with fewer resources. Despite previous efforts by school officials to avoid classroom-related cuts, it appears likely that the current fiscal crunch will have a direct impact on instructional services and supplies. City school systems, for example, are proposing to balance their budgets through increased class sizes, employee furloughs, shortened school years, and freezes on textbook and furniture purchases (Gewertz and Reid 2003). These sorts of strategies to accommodate shrinking budgets have the potential to severely reduce the likelihood that these school systems will make progress toward meeting the increasingly high performance standards being set by states.

The current education policy emphasis on higher performance standards, school-level accountability, and market-oriented reform presents important research challenges within the field of school finance and the economics of education. The simultaneous pursuit of both equity and efficiency within this policy context creates an unprecedented demand for rigorous, timely, and field-relevant research on fiscal practices in schools.

In an effort to help meet this demand, we have developed a new book series, *Research in Education Fiscal Policy and Practice*. For our inaugural volume, *Fiscal*

[1] Gewertz, C., and Reid, K.S. (2003, February 5). Hard Choices: City Districts Making Cuts. *Education Week, 22*(21), 1.

Policy in Urban Education (Roellke and Rice 2002), we assembled a group of school finance experts to address the critical challenges in urban education fiscal policy and practice. These researchers contributed a diverse set of policy papers, including analyses of fiscal accountability in urban schools, teacher recruitment and quality, private influences on urban school funding, and other pressing concerns in urban education reform.[2] While many of the issues addressed by the analyses are applicable to all schools—urban, suburban, and rural—we intentionally elected to focus on questions of both policy and practice that impact schools located in the nation's inner cities. Our intent here is to synthesize the key themes that surfaced from this research effort and to highlight compelling issues in need of further examination.[3] While by no means exhaustive, the sections below outline some of the most current and critical issues facing urban fiscal policy.

The Complexity of Urban School Reform

The chapters in our volume make it very clear that education reform in urban contexts is extremely complex, even in the best of times. For decades, urban schools have been the object of one intervention after another, exemplary of Cuban's phrase, "reforming again, again, and again" (1990). Today's policy climate is characterized by widespread intolerance toward schools with a history of failure. Low-performing schools, often located in urban areas that serve large numbers of poor and minority children, report chronically low attendance, achievement, and graduation rates. Many of these schools have been the focus of persistent reform efforts, yet satisfactory levels of student performance have remained elusive. Moreover, even when progress is evident, it is often not at a pace to keep up with state demands to achieve higher standards as measured by state-mandated tests (Corbett and Wilson 1991; Ladd 1996).

Education reform in urban contexts is extremely complex, even in the best of times.

The educational policy terrain continues to be flooded with options like class-size reduction, alternative scheduling, summer enrichment, early intervention programs, as well as a wide array of whole-school reform models. Successful implementation of these initiatives is highly dependent on a variety of contextual factors, including student demographics, fiscal capacity, school size, spending level, and district/school governance (Iatarola, Stiefel, and Schwartz 2002; Brent, Roellke, and Monk 1997). This is certainly the case in all schools, but is especially problematic for urban educators, who confront a high concentration of at-risk students and a wide diversity of competing reforms that focus on at-risk youth. Further, researchers offer little definitive evidence on the cost or the effectiveness of alternative investment options, information that could be very helpful to local policymakers facing a multitude of policy alternatives and a limited stock of resources.[4]

While many school reform strategies have been targeted to urban schools, a great deal of attention in education reform circles has focused on comprehensive school reform models as promising alternatives for improving the effectiveness of underperforming schools serving large concentrations of at-risk students. This approach to reform is attractive in that each model prescribes a "configuration of resources" that are intended to have a positive effect on the entire educational experience of students during their elementary school years (Rice 2001). Policies encouraging schools to adopt comprehensive school reform

[2] For additional details of these research studies, see Roellke and Rice (2002).

[3] This paper draws from the final summary chapter of the volume, Rice and Roellke (2002), and from Rice and Roellke (2003).

[4] A series of literature reviews by Hanushek (1981, 1986, 1996, 1997) have shown a high level of inconsistent and insignificant findings across studies estimating the impact of various types of educational investments. On the other hand, researchers who have reanalyzed Hanushek's data, challenging both his assumptions and his basic "vote counting" methodology, have reported more positive and consistent interpretations of the same set of studies (Hedges, Laine, and Greenwald 1994; Laine, Greenwald, and Hedges 1996; Krueger 2002). On the cost side, Levin and McEwan (2001) and Rice (1997, 2001) argue that cost analysis is an underutilized analytic tool in the field of education.

models are evident at the federal, state, and local levels, and much of this attention has focused on urban schools and school systems.

Research that examines both the implementation process and achievement effects of these comprehensive reform strategies has yielded mixed results.[5] Methodological challenges, including contextual differences across model sites and the lack of randomized experiments, make it difficult to draw definitive conclusions. In addition, these comprehensive reforms must be studied longitudinally with a specific need for careful cost-effectiveness studies (Bifulco 2002).

Two chapters in our volume on fiscal policy issues in urban education address key issues related to comprehensive school reform. Bifulco reviews what the existing evidence shows to draw conclusions about the degree to which whole-school reform models can be expected to enhance the productivity of schools. His focus is on three models that have been implemented and studied: (1) The School Development Program, (2) Success for All, and (3) Accelerated Schools. In terms of effectiveness, he finds some evidence of positive effects of the Success for All models, but recognizes that the studies reporting the strongest positive effects were conducted by program developers. Bifulco concludes, "The School Development Program does not appear to have positive impacts on average. However, it may have positive impacts in some places under certain conditions" (p. 29). Finally, the evidence on Accelerated Schools, Bilfulco concludes, is insufficient to draw general conclusions about the program's effectiveness.

In terms of the costs, Bifulco provides a summary table (table 1) of cost estimates generated by a handful of researchers who have studied the costs of these three models. While the range of costs for any given model is substantial, the Success for All program is generally the most resource-intensive of the three approaches. This is due to the highly prescriptive nature of this model in terms of new positions and training requirements. Bifulco concludes that "the relatively large amount of resources demanded by Success for All raises questions about whether its positive impacts are the result of increased productivity at SFA schools or merely increased resources. The answer to this question hinges on whether the resources used to implement SFA represent additional resources or a reallocation of existing resources" (p. 30).

Bifulco concludes that "no model can improve student outcomes in all schools . . . More research is needed to identify the conditions under which particular whole-school reform models are most likely to succeed" (p. 30). He recommends four directions for future research on whole-school reform models: evaluations of more models, additional cost studies, estimates of long-term impacts, and efforts to identify factors that promote the success of whole-school reform.

[5] For a review of the evidence of comprehensive school reform models on student achievement, see Herman (1999).

Table 1. Estimates of costs for three whole-school reform models

Study	Success for All (in dollars)	School Development Program (in dollars)	Accelerated Schools (in dollars)
King (1994)[1]	261,060–646,500	102,800–278,150	48,000–266,000
Barnett (1996)[2]	160,500–340,500	57,500–219,000	17,000–80,000
Herman (1999)[3]	270,000	45,000	27,000
Borman and Hewes (2001)[4]	153,293–578,550	—	—

—Not available.

[1] Estimates assume school with 500 students, and do not include costs for materials.

[2] Estimates assume school with 500 students, and do not include costs of parental time.

[3] Estimates are based, in part, on King (1994) and Barnett (1996).

[4] Estimates for actual schools of varying sizes. Estimates do not include extra time devoted by existing staff for training and implementation activities.

SOURCE: Bifulco (2002).

Another chapter in the volume underlines the idea that successful adoption of whole-school reform requires a careful, inclusive model selection process and ongoing support and guidance from model developers (Erlichson and Goertz 2002; Hertling 1999). Erlichson and Goertz report findings from a study examining 3 years of implementation of whole-school reform and school-based budgeting in New Jersey. The authors examined the response of schools and school districts to *Abbott v. Burke*, the May 1998 New Jersey Supreme Court decision that directed schools in 30 poor urban districts to adopt comprehensive school reform programs. "In practical terms, the Court's mandate required nearly 450 schools in New Jersey to adopt such models in a three-year period as the primary way of addressing the special education needs of urban students" (p. 37). The Court also required that each school adopt school-based budgeting concurrently with the comprehensive school reform model.

One feature of this decision was that schools could choose among a variety of reform models. Erlichson and Goertz provide a breakdown of which models were selected and when they were implemented during the 3-year period (table 2).

Based on data collected through site visits, interviews, and questionnaires, Erlichson and Goertz draw a number of conclusions regarding the implementation of these court-ordered reforms. Regarding the implementation of whole-school reform models, the authors identified six shortcomings that undermined the process:

(1) Flawed model selection process—particularly inadequate information regarding the different options

(2) Mismatch between expectations of a model and reality—a direct result of the flawed model selection process

(3) Absence of a link to core curriculum content standards—standards that were established by the state and for which schools are held accountable

(4) Lack of time—to train, plan, and establish a new vision to implement the model

(5) Lack of consistent and meaningful support from developers—resulting, in part, from insufficient model staff to accommodate the influx of New Jersey schools

Table 2. Statewide implementation of comprehensive whole-school reform models: By model and cohort (year of implementation)

Model	Cohort: (Number of schools implementing each model is reported for each cohort) One	Two	Midyear Second	Three	Midyear Third	Total
Year of implementation	*1998–1999*	*1999–2000*	*2000–2001*	*2000–2001*	*2001–2002*	
Accelerated Schools Program	1	14	10	7	0	32
America's Choice	0	6	3	8	4	21
Atlas	0	0	0	1	0	1
Coalition of Essential Schools	3	3	8	32	13	59
Community for Learning	23	8	3	2	0	36
Co-Nect	0	7	4	19	3	33
High Schools That Work	0	0	0	2	6	8
Microsociety	0	0	1	0	0	1
Modern Red Schoolhouse	2	5	0	2	0	9
Paideia	0	1	0	2	1	4
School Development Program (Comer)	16	13	7	76	6	118
Success for All/Roots and Wings	27	23	5	13	1	69
Talent Development	0	1	0	1	9	11
Ventures	0	2	0	11	6	19
Alternative Program Design (approved)	0	0	0	6	7	13
Total whole-school reform schools	72	83	41	182	56	434

SOURCE: Erlichson and Goertz (2002).

(6) Lack of meaningful and consistent support from New Jersey Department of Education field staff—a team was assigned to every school implementing whole-school reform

In the end, the authors conclude with three lessons from the New Jersey case. First, the success of mandated school reform depends on both the will and skill of educators across the levels of the system, but particularly in the targeted schools. Second, the successful implementation of reform can be undermined by complex governance structures, unclear roles, and conflicting messages. Finally, the multiple actors in the education system can limit the degree of school empowerment that is realized, even in cases when the reform is designed to give schools greater discretion in how they approach reform.

Teacher Supply and Quality in Urban Schools

High-quality teachers are a fundamental resource needed to realize the high standards characteristic of most state accountability plans. While researchers have debated the extent of the teacher supply problem nationally, there is general agreement that schools serving large numbers of poor and minority students face the greatest challenges in recruiting and retaining a faculty of qualified teachers. Further, urban schools tend to serve high concentrations of students with these characteristics, making teacher quality issues a key concern in urban education.

Schools serving large numbers of poor and minority students face the greatest challenges in recruiting and retaining a faculty of qualified teachers.

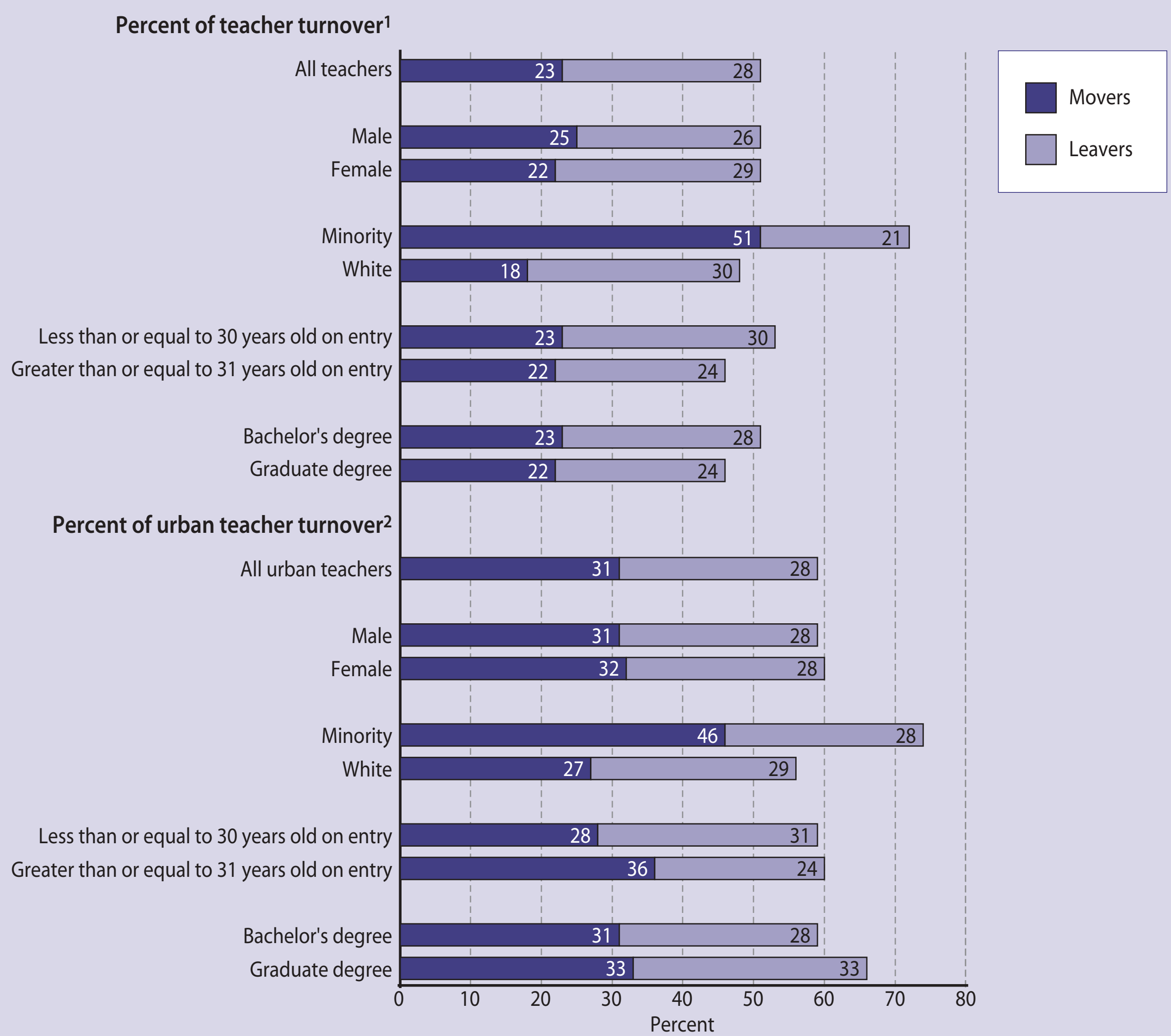
While recruiting high-quality teachers to urban, difficult-to-staff schools is a challenge, retaining those teachers over time is just as critical (Allgood and Rice 2002). Recent research on the New York teaching workforce, for example, found that quit rates in the state are highest in New York City. In addition to these high quit rates, evidence suggests that teachers who leave New York City schools generally possess better qualifications than those who remain (Lankford, Loeb, and Wyckoff 2002). These attrition patterns are consistent with earlier claims made that one out of every five New York City teachers leave after the first year and one of every three teachers leave after 3 years (Schwartz 1996). Resources aimed at elevating teacher quality in urban schools should be targeted not only toward drawing high-quality teachers to difficult-to-staff schools, but also toward reducing teacher turnover in those schools.

Two chapters in our volume on fiscal policy issues in urban education address the issue of teacher attrition in urban districts. Theobald and Michael examine teacher attrition among novice teachers over 5 years in four midwestern states (Ilinois, Indiana, Minnesota, Wisconsin). They examine four categories of novice teachers, "those who: (a) taught continuously in the same district all 5 years ('stayers'), (b) transferred to another school district(s) in the state, but remained in the state for all 5 years ('movers'), (c) left public school teaching in a state and did not return ('leavers'), and (d) left public school teaching in a state, but returned ('returnees')." Figure 1 shows the percentage of teacher turnover in 5 years among the 11,787 teachers who entered the profession in the four states in 1995–96. As can be seen from this figure, leaving teaching is related to certain personal characteristics such as race, age, and level of education.

Figure 1 also provides similar information about the 3,194 novice urban teachers who entered the profession in the 1995–96 school year. These data reveal that "urban teachers—regardless of their gender, race, age, or degree status—are significantly more likely to move out of their district than are novice teachers hired by non-urban districts" (Theobold and Michael 2002, p. 144). The findings here underline the importance of including movers in this sort of analysis, since the results for leavers alone would lead to conclusions of little difference between novice teachers hired by urban districts and those hired by nonurban districts.

Theobald and Michael also present interesting findings regarding teacher attrition among novice teachers by level of education and subject area. Figure 2 reports these findings for all novice teachers hired during the 1995–96 academic year, and for novice urban teachers hired during that school year. Figure 2 shows that, among all novice teachers, mathematics teachers and science teachers (except for biology) are more likely to leave teaching and are less likely to transfer among school districts. This finding could be explained by the alternatives available to these individuals in the broader labor market. The urban data presented in figure 2 reveal that urban teachers in special education, business, foreign language, and mathematics are more likely to move to other districts than their counterparts who were hired in nonurban districts. Again, a focus only on leavers would lead to a conclusion of little difference in teacher turnover between urban and nonurban schools.

Figure 1. Percentage of teacher turnover and percentage of urban teacher turnover in 5 years among teachers who entered the profession in four midwestern states in 1995–96: By selected personal characteristics

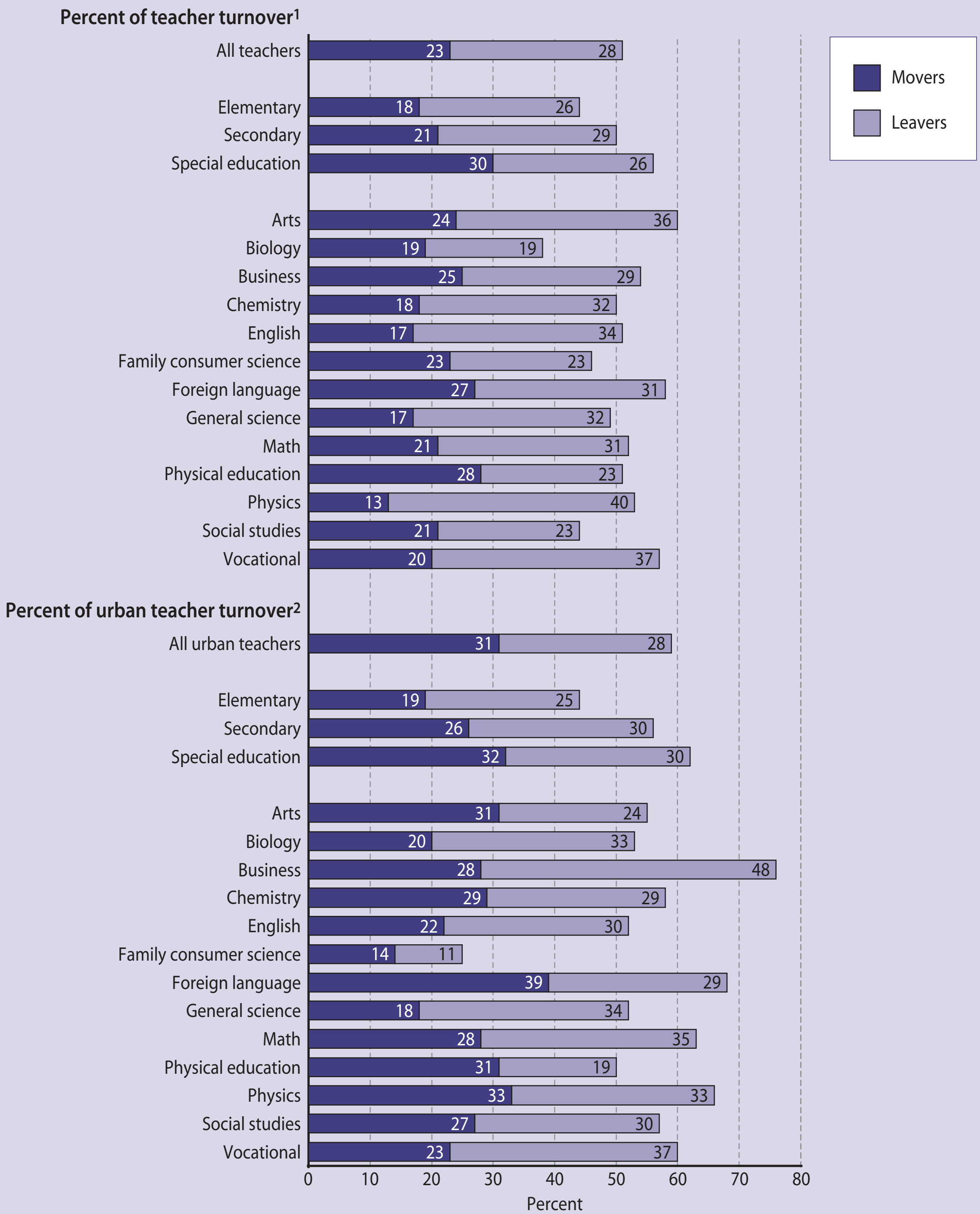


[1] Among 11,787 teachers.

[2] Among 3,194 teachers.

SOURCE: Theobald and Michael (2002).

Figure 2. Percentage of teacher turnover and percentage of urban teacher turnover in 5 years among teachers who entered the profession in four midwestern states in 1995–96: By selected professional characteristics



[1] Among 11,787 teachers.

[2] Among 3,194 teachers.

SOURCE: Theobald and Michael (2002).

Theobald and Michael conclude their chapter with a set of policy recommendations. These include providing more funding to school districts serving disproportionately large concentrations of disadvantaged students, creating an external context in all school districts that is supportive of novice teachers, and providing pay premiums for novice teachers in districts that face high turnover rates.

In another chapter, Jennifer Imazeki examines teacher attrition in Wisconsin, paying particular attention to the role of wage differentials. Her data suggest that teachers do respond to wages; in other words, increasing teacher salaries can help to slow attrition. She also addresses the question, How much do salaries need to increase? To answer this question, Imazeki simulates several wage-increase scenarios involving a $5,000 salary increase. Table 3 presents her results for Milwaukee. Imazeki points out that even without the $5,000 wage differential, Milwaukee teacher salaries are higher than those elsewhere in the state (ranging from one-half of a standard deviation for beginning teachers to one and a half standard deviations for maximum salaries).

When $5,000 is offered in addition to current salaries, the effect depends on the scenario. Exits are most responsive to an increase for all novice teachers, while transfers are responsive to changes in relative salary (i.e., targeted to Milwaukee). In general, for both men and women, $5,000 is not enough to bring overall attrition rates in Milwaukee to the levels of the average district in Wisconsin. While transfer rates reach the level of the average district, exit rates remain higher in Milwaukee.

One explanation for the inability of financial incentives to solve the problem is that challenging working conditions further complicate the low supply and high attrition rates of urban teachers. Urban teachers, for example, educate a disproportionate number of students with special needs and confront the highest percentages of students who are not proficient in English. Urban teachers are much more likely than suburban and rural teachers to report problems such as high absenteeism, serious student violence, and poor parental involvement (Lippman, Burns, and McArthur 1996; Imazeki 2002; Van Horn 1999). As a result, compensation-based strategies alone are not likely to be sufficient to attract and retain high-quality teachers in urban schools. Policies must attend to the working conditions in these schools, in addition to providing targeted salary incentives, if they are to realize a high-quality urban teaching workforce (Allgood and Rice 2002).

Addressing concerns about the teacher workplace and professional climate requires moving beyond state-level and even district-level reform strategies. Because most recruitment and induction activities occur within local schools and classrooms, it is important that policymakers at more centralized levels of the system be attentive to varying local capacities. Solutions to the teacher supply and quality problem in our inner cities may require a more concerted effort to support creative, locally designed strategies to enhance the professional environment of emerging educators (Roellke and Meyer 2003; Theobald and Michael 2002).

Table 3: Survivor functions for Milwaukee: Simulations with $5,000 added to salaries

		A		B		C		D		E	
	Duration of teaching spell (years)	Actual Salary		$5,000 for all beginning teachers		$5,000 for beginning teachers in Milwaukee		$5,000 for all experienced teachers		$5,000 for all teachers in Milwaukee	
		Women	Men	Women	Men	Women	Men	Women	Men	Women	Men
Exits											
	1	81.1	83.6	83.7	85.9	82.4	83.2	81.6	85.5	82.9	85.1
	3	59.4	66.8	63.9	70.7	61.6	65.8	60.4	70.1	62.5	69.3
	5	50.1	56.6	54.9	61.2	52.3	55.4	51.1	60.6	53.3	59.4
Transfers											
	1	97.2	94.9	97.3	95.9	97.9	96.2	97.4	95.1	98.1	96.3
	3	93.1	88.1	93.3	90.2	94.9	90.9	93.7	88.6	95.3	91.3
	5	89.5	84.3	89.8	86.9	92.1	87.9	90.4	84.8	92.8	88.3

SOURCE: Imazeki (2002).

Nontraditional Support for Urban Schools

Most analyses of fiscal education policy tend to focus on revenues generated by local, state, and federal governments. However, school leaders and policymakers are taking advantage of a broader resource base than is typically recognized in more traditional fiscal analyses. Fiscal and personnel support for students in urban schools is also derived from nontraditional sources, including private foundations, volunteer networks, and other human service agencies. Two chapters in our volume address these nontraditional resources, which can provide for a variety of urban school services, including tutoring, vocational counseling, literacy programs, and even teacher training (Schwartz, Amor, and Fruchter 2002). Schwartz and her colleagues use a combination of school budget data and survey data to find that the vast majority of public schools receive some sort of nontraditional or private support, or both. Table 4 illustrates that for New York City the amount of such support varies significantly across schools, reflecting the interplay of donor preferences, the ability of school leaders to solicit funds, and the political economy on nonentitlement government funding.

In some cases, these nontraditional resources can account for over half of the financing of children's services within urban school districts (Picus et al. 2002). Picus and his colleagues take into consideration not only school budgets, but also other public spending associated with children's services. Table 5 provides both low and high estimates of this spending for health and social services in the University of Southern California area.

These findings imply that private and nontraditional resources devoted to education and other children's services are not trivial, and therefore should be more common in fiscal analyses of education. Further, the research indicates a substantial level of variability in the amount and the distribution of nontraditional resources, which has clear implications for both equity and efficiency of urban schools and school systems. Finally, the capacity of schools to adopt certain reforms

Table 4. Nontraditional revenues in New York City public schools (N = 1,023 schools)

Variable	Mean	Minimum	Maximum
Private support per pupil (in dollars)	59.0	1.0	1,901.7
State grants per pupil (in dollars)	104.3	8.4	2,402.7
Federal grants per pupil (in dollars)	52.5	1.9	820.3
Other grants per pupil (in dollars)	132.9	68.5	386.3
Total per pupil (in dollars)	348.4	104.9	2,873.6
Total spending per pupil (in dollars)	8,245.7	1,673.1	22,414.4
Percent White	15.7	0.0	93.8
Percent Black	36.7	0.2	97.6
Percent Hispanic	37.5	1.3	98.1
Percent Asian	10.2	0.0	94.3
Percent female	49.1	6.0	87.9
Percent immigrant	8.2	0.0	96.3
Percent limited English proficient (LEP)	15.9	0.1	100.0
Percent free lunch	71.7	5.9	100.0
Percent special education	6.1	0.0	37.7
Enrollment	1,001.5	51.0	5,004.0

NOTE: *Private support per pupil* includes contributions of cash, equipment, and services. *State grants per pupil* include legislative grants, magnet grants, and Comprehensive Instructional Management System grants. *Federal grants per pupil* include magnet grants and federal bilingual program (Title 7) grants. *Other (non-entitlement) grants per pupil* include capital projects, building Board of Education/Office for Development maintenance, student information services, early grade paraprofessionals redeployment, self-sustaining accounts, Employment Prep Education Program, city-funded programs, and food services. The total is the sum of the above. *Total spending per pupil* includes direct services to schools, district and superintendency costs, and systemwide costs (no pass-throughs). Direct services to schools include classroom instruction, instructional support services, school leadership, ancillary support services, building services, and district support; district/superintendency costs include instructional support/administration and other district/borough costs; systemwide costs include central instructional support, central administration, and other obligations.

SOURCE: Schwartz, Bel Hadj Amor, and Fruchter (2002).

may depend on the availability of nontraditional sources of support. For instance, since several whole-school reform designs rely heavily on the use of volunteers and community resources, policymakers must be attentive to the availability (or lack thereof) of this important pool of nontraditional and private support.

As market-oriented reforms gain momentum, it is clear that additional attention to these nontraditional revenues and data on private school finance are needed. The choice provisions contained within the federal No Child Left Behind Act illustrate this ongoing interest in experimenting with market-based mechanisms as a means for reforming public education. This is also evident in the privatization of public schools in several urban districts like Philadelphia and Baltimore, the expansion of charter schools throughout the country, and the increasing attention being given to vouchers as a means to allow students enrolled in failing public schools to elect to attend private schools. We know very little, for example, about the manner in which private schools secure, allocate, and use educational resources (Brent 2002).

Accountability and Adequacy in Urban Schools

The current policy climate characterized largely by high-stakes accountability introduces a variety of monitoring and assessment issues. These issues are especially salient for urban schools where student performance is often a concern. While much of the attention associated with monitoring school systems has focused on testing, broader accountability goals require more sophisticated analyses that assess fiscal condition and capacity, opportunity to learn, and other equity-related issues. It is clear that urban districts struggle to meet the demands of fiscal accountability and increased standards for student achievement (Alexander 2002).

This gap between higher achievement standards and the resources required to reach them has resulted in a series of legal challenges focused on adequacy. Often referred to as the "third wave" of school finance litigation, plaintiffs argue that finance formulas prevent poor school districts from providing an adequate education as defined by state education clauses.[6] An important and recent third wave victory for plaintiffs is the New York Court of Appeals ruling in *Campaign for Fiscal Equity v. State of New York* (2003). The 4-1 decision overturned a 2002 state appellate court ruling that the state was responsible for providing only an eighth- or ninth-grade education. The higher court ruled that a "sound basic education" goes beyond eighth- or ninth-grade, and should include a "meaningful high school education." The remedy laid out by the Court of Appeals requires a costing-out study (to be completed by March 2004) to determine a

[6] For a more detailed discussion of school finance litigation, see Roellke, Green, and Zielewski (in press).

Table 5. Total estimated health and social service expenditures in the University of Southern California area

Service provider	Amount (in dollars)	Percent of total
Low estimate		
County of Los Angeles	223,833,900	55.16
City of Los Angeles	15,403,054	3.80
Not-for-profit agencies	48,766,800	12.02
Los Angeles Unified School District	117,818,860	29.03
Total	405,822,614	
High estimate		
County of Los Angeles	223,833,900	40.77
City of Los Angeles	15,403,054	2.81
Not-for-profit agencies	48,766,800	8.88
Los Angeles Unified School District	260,994,110	47.54
Total	548,997,864	

SOURCE: Picus, McCroskey, Robillard, Yoo, and Marsenich (2002).

dollar amount that can ensure that all students have the opportunity to obtain this higher level of achievement specified by the Court.

Concluding Remarks

The pressure for urban schools to operate efficiently, equitably, and adequately is unprecedented. As we indicated in our introduction, this challenge has become even more daunting in light of dramatic budget cuts for urban schools. Solutions to the problems of urban schools are not easily answered by research. We are confident, nonetheless, that emerging studies on whole-school reform, teacher supply and quality, nontraditional revenues, high-stakes accountability, and other pressing fiscal policy issues can assist policymakers and school leaders in their quest for increased equity and enhanced productivity in urban schools. The second volume of our series, guest edited by Faith Crampton and David Thompson, focuses on school infrastructure funding in both the United States and Canada.[7] A diverse set of school finance and policy analyses are included in this second volume, including capital needs in urban and rural schools; specific infrastructure considerations for students with disabilities; school finance litigation as a strategy for improving school facilities; and the role of school administrators in school renovation projects. Our goal is that these volumes, along with others that follow in the series, can assist academic researchers, policymakers, and school practitioners in their efforts to improve education fiscal policy and practice.

7 See Crampton and Thompson (2003).

References

Alexander, N. (2002). Balancing the Books: The Relationship Between Fiscal Condition, Educational Output, and District Competition in the Metropolitan Districts of Minnesota. In C.F. Roellke and J.K. Rice (Eds.), *Fiscal Policy in Urban Education* (pp. 65–92). Greenwich, CT: Information Age.

Allgood, W., and Rice, J.K. (2002). The Adequacy of Urban Education: Focusing on Teacher Quality. In C.F. Roellke and J.K. Rice (Eds.), *Fiscal Policy in Urban Education* (pp. 155–180). Greenwich, CT: Information Age.

Barnett, W.S. (1996). Economics of School Reform: Three Promising Models. In H.F. Ladd (Ed.), *Holding Schools Accountable: Performance-Based Reform in Education.* Washington, DC: The Brookings Institution.

Bifulco, R. (2002). Can Whole-School Reform Improve the Productivity of Urban Schools? The Evidence on Three Models. In C.F. Roellke and J.K. Rice (Eds.), *Fiscal Policy in Urban Education* (pp. 11–35). Greenwich, CT: Information Age.

Borman, G.D., and Hewes, G.M. (2001). *The Long-Term Effects and Cost-Effectiveness of Success for All.* Baltimore, MD: Johns Hopkins University, Center on Research for the Education of Students Placed at Risk.

Brent, B.O. (2002). Private School Finance: Tugging on the Curtain. In C.F. Roellke and J.K. Rice (Eds.), *Fiscal Policy in Urban Education* (pp. 205–229). Greenwich, CT: Information Age.

Brent, B.O., Roellke, C.F., and Monk, D. (1997). Understanding Teacher Resource Allocation in New York's Secondary Schools: A Case Study Approach. *Journal of Education Finance, 23*(2), 207–233.

Campaign for Fiscal Equity v. State of New York, 100 N.Y.2d 893, 2003 slip op. 15615 (N.Y. June 26, 2003).

Corbett, H.D., and Wilson, B.L. (1991). Testing, Reform, and Rebellion. Norwood, NJ: Ablex.

Crampton, F.E., and Thompson, D.C. (Eds.). (2003). *Saving America's School Infrastructure*. Greenwich, CT: Information Age.

Cuban, L. (1990). Reforming Again, Again, and Again. *Educational Researcher, 19*(1), 3–13.

Erlichson, B.A., and Goertz, M. (2002). Whole School Reform and School-Based Budgeting in New Jersey: Three Years of Implementation. In C.F. Roellke and J.K. Rice (Eds.), *Fiscal Policy in Urban Education* (pp. 37–64). Greenwich, CT: Information Age.

Gewertz, C., and Reid, K.S. (2003, February 5). Hard Choices: City Districts Making Cuts. *Education Week, 22*(21), 1.

Hanushek, E.A. (1981). Throwing Money at Schools. *Journal of Policy Analysis and Management, 1*(1), 19–41.

Hanushek, E.A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature, 24*(3), 1141–1177.

Hanushek, E.A. (1996). School Resources and Student Performance. In G. Burtless (Ed.), *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success* (pp. 43–73). Washington, DC: Brookings Institution Press.

Hanushek, E.A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis, 100*(1): 84–117.

Hedges, L.V., Laine, R.D., and Greenwald, R. (1994). Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes. *Educational Researcher, 23*(3), 5–14.

Herman, R. (1999). *An Educators' Guide to Schoolwide Reform*. Arlington, VA: Educational Research Service. (ERIC ED460429)

Hertling, E. (1999). *Implementing Whole-School Reform* (ERIC Digest No. 128). Eugene, OR: ERIC Clearinghouse on Educational Management. (ERIC ED454565)

Iatarola, P., Stiefel, L., and Schwartz, A.E. (2002). School Performance and Resource Use: The Role of New York City Public School Districts. In C.F. Roellke and J.K. Rice (Eds.), *Fiscal Policy in Urban Education* (pp. 93–116). Greenwich, CT: Information Age.

Imazeki, J. (2002). Teacher Attrition and Mobility in Urban Districts: Evidence From Wisconsin. In C.F. Roellke and J.K. Rice (Eds.), *Fiscal Policy in Urban Education* (pp. 119–136). Greenwich, CT: Information Age.

King, J.A. (1994). Meeting the Educational Needs of At-Risk Students: A Cost Analysis of Three Models. *Educational Evaluation and Policy Analysis*, *16*(1): 1–19.

Krueger, A.B. (2002). Understanding the Magnitude and Effect of Class Size on Student Achievement. In L. Mishel and R. Rothstein (Eds.), *The Class Size Policy Debate* (pp. 7–36). Washington, DC: Economic Policy Institute.

Ladd, H. (1996). *Holding Schools Accountable: Performance-Based Reform in Education*. Washington, DC: Brookings Institution Press.

Laine, R.D., Greenwald, R., and Hedges, L.V. (1996). Money Does Matter: A Research Synthesis of a New Universe of Education Production Function Studies. In L.O. Picus and J.L. Wattenbarger (Eds.), *Where Does the Money Go? Resource Allocation in Elementary and Secondary Schools* (pp. 44–70). Thousand Oaks, CA: Corwin.

Lankford, H., Loeb, S., and Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation and Policy Analysis 24*(1): 37–62.

Levin, H.M., and McEwan, P. (2001). *Cost-Effectiveness Analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Lippman, L., Burns, S., and McArthur, E. (1996). *Urban Schools: The Challenge of Location and Poverty* (NCES 96–184). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Picus, L.O., McCroskey, J., Robillard, E., Yoo, J., and Marsenich, L. (2002). Using Student Level Data to Measure School Finance Adequacy: An Exploratory Analysis. In C.F. Roellke and J.K. Rice (Eds.), *Fiscal Policy in Urban Education* (pp. 181–201). Greenwich, CT: Information Age.

Rice, J.K. (1997). Cost Analysis in Education: Paradox and Possibility. *Educational Evaluation and Policy Analysis, 19*(4), 309–317.

Rice, J.K. (2001). Illuminating the Black Box: The Evolving Role of Education Productivity Research. In S. Chaikind and W. Fowler (Eds.), *Education Finance in the New Millennium, 2001 Yearbook of the American Education Finance Association* (pp. 121–138). Larchmont, NY: Eye on Education.

Rice, J.K., and Roellke, C.F. (2002). Fiscal Policy in Urban Education: Lessons Learned and Implications for Research. In C.F. Roellke and J.K. Rice (Eds.), *Fiscal Policy in Urban Education* (pp. 253–257). Greenwich, CT: Information Age.

Rice, J.K., and Roellke, C.F. (2003). Urban School Finance: Increased Standards and Accountability in Uncertain Economic Times. *School Business Affairs, 69*(5), 30–33.

Roellke, C.F., Green, P., and Zielewski, E.H. (in press). School Finance Litigation and Legislation: The Unfulfilled Promise of the Third Wave. Invited paper being prepared for a special issue on school finance. *Peabody Journal of Education*.

Roellke, C.F., and Meyer, T. (2003). Recruitment, Induction, and Retention of “Academically Talented” Urban School Teachers: Evidence From New York City. In M. Plecki and D. Monk (Eds.), *School Finance and Teacher Quality: Exploring the Connections*. Larchmont, NY: Eye on Education.

Roellke, C.F., and Rice, J.K. (2002). *Fiscal Policy in Urban Education*. Greenwich, CT: Information Age.

Schwartz, A.E., Bel Hadj Amor, H., and Fruchter, N. (2002). Private Money/Public Schools: Early Evidence on Private and Non-Traditional Support for New York City Public Schools. In C.F. Roellke and J.K. Rice (Eds.), *Fiscal Policy in Urban Education* (pp. 231–250). Greenwich, CT: Information Age.

Schwartz, F. (1996). Why Many New Teachers Are Unprepared to Teach in Most New York City Schools. *Phi Delta Kappan 78*(1): 82–84.

Theobald, N.D., and Michael, R.S. (2002). Reducing Novice Teacher Attrition in Urban Districts: Focusing on the Moving Target. In Roellke, C.F., and Rice, J.K. (Eds.), *Fiscal Policy in Urban Education* (pp. 137–152). Greenwich, CT: Information Age.

Van Horn, R. (1999). Inner-City Schools: A Multi-Variable Discussion. *Phi Delta Kappan 81*(4), 291–297.

Financing Education So That No Child Is Left Behind: Determining the Costs of Improving Student Performance

Andrew Reschovsky
University of Wisconsin-Madison

Jennifer Imazeki
San Diego State University

About the Authors

Andrew Reschovsky is a professor of applied economics and public affairs at the University of Wisconsin-Madison. He is a public finance economist with a Ph.D. in economics from the University of Pennsylvania. He has conducted research on a wide range of public finance and tax policy issues. His recent work includes work on determining the costs of educational adequacy and on the financing of rural education. He has just returned from South Africa, where he has been advising the Government of South Africa on the design of education funding formulas and on a number of education finance issues involving both early childhood education and adult basic education. He has published numerous articles in professional journals and books, and has written on public issues for newspapers and magazines. He can be contacted at reschovsky@lafollette.wisc.edu.

Jennifer Imazeki is an associate professor of economics at San Diego State University, where she teaches courses in public finance and applied microeconomics. She received her Ph.D. in economics from the University of Wisconsin-Madison, where she also worked as a researcher for the Consortium for Policy Research in Education. Her research focuses on school finance reform and teacher labor markets. Recent work includes papers on voluntary contributions to public schools and on the costs of rural education. She can be contacted at jimazeki@mail.sdsu.edu.

Financing Education So That No Child Is Left Behind: Determining the Costs of Improving Student Performance

Andrew Reschovsky
University of Wisconsin-Madison

Jennifer Imazeki
San Diego State University

Acknowledgments

This paper is a slightly revised version of the authors' article, "Let No Child Be Left Behind: Determining the Cost of Improving Student Performance," published in *Public Finance Review* (vol. 31, no. 3, May 2003, pp. 263–290), copyright © 2003 Sage Publications, and reprinted by permission of Sage Publications, Inc.

Financial support for the research was provided by the Consortium for Policy Research in Education (CPRE) (Grant No. OERI-R3086A60003) and by the Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison. The opinions expressed in this article are those of the authors and do not reflect the views of the National Institute on Educational Governance, Finance, Policy-Making and Management, Institute of Education Sciences (formerly Office of Educational Research and Improvement), U.S. Department of Education; the institutional partners of CPRE; or the Wisconsin Center for Education Research.

Introduction

Improving the quality of primary and secondary education is a top priority of the Bush administration and of members of both parties in Congress. In introducing his education proposals to Congress, the president highlighted the "academic achievement gap" that exists between students from rich and poor families and between White and minority children (U.S. Department of Education 2001). As evidence of the failure of the current system, the president cited the fact that "nearly 70 percent of inner city fourth graders are unable to read at a basic level on national reading tests." The primary goal of the administration's education policy is to close this achievement gap and to ensure that, in President Bush's oft-repeated phrase, "no child should be left behind."

To that end, the Bush administration proposed, and after extensive debate, Congress enacted, the No Child Left Behind Act of 2001. The new legislation mandates annual testing of all students in grades 3 through 8 and requires that schools make annual progress in meeting student performance goals for all students and for separate groups of students characterized by race, ethnicity, poverty, disability, and limited English proficiency (U.S. Department of Education 2002). The underlying premise of the legislation is that schools must be held accountable for the academic performance of their students. The legislation will reward schools that succeed in meeting state-imposed achievement goals and will sanction schools that fail. The intent is that all students, but especially students from disadvantaged backgrounds, show annual improvements in their academic performance as measured against state standards. Measuring student performance is thus a necessary component in a policy designed to improve the quality of education. We doubt it is a sufficient policy. In this article, we present evidence suggesting that measuring student performance, setting performance standards, and threatening to sanction schools that fail to meet these standards are unlikely to close achievement gaps unless accompanied by a restructuring of the financing of public education. It should be noted that although the new federal legislation represents the first time that all states will be required to test students on an annual basis, some states have been administering such tests for a number of years. Despite this testing and the publication of the results on an individual school basis, the performance of students in many schools, especially those serving disadvantaged children, remains substantially below average.

Setting performance standards is unlikely to close achievement gaps unless accompanied by a restructuring of the financing of public education.

We suggest that the amount of money necessary to meet student performance standards will vary across school districts. This variation in costs will be due primarily to factors over which local school officials have little control. For example, a school district with a high concentration of students from poor families or from families where English is not spoken in the home may have to use additional resources (in the form of smaller classes or specialized programs) to reach specified achievement goals. Also, some districts, given their location and the composition of their student bodies, will have to pay higher salaries than other districts to attract high-quality teachers.

Requiring that all schools increase the academic performance of their students is a potentially important step in improving the quality of education in the United States. However, if cost differences among school districts are substantial, then imposing statewide student performance standards without simultaneously allocating more state financial aid to school districts with high costs may result in a situation in which school districts with above-average costs will not have enough resources to educate their students to meet the new standards. These schools may fail, not necessarily because of their own inability to effectively educate children but because they have insufficient fiscal resources to do the job.

Although the new federal education legislation does not explicitly address the connection between the cost of education and student performance, over the past decade the courts in a number of states have explicitly recognized the link between educational finance and student performance. In several states, the courts have declared state school financing systems unconstitutional because they have failed to provide all students, and especially those from economically disadvantaged families, with a sufficiently high-quality education; in the language favored by the courts, these systems have failed to provide an adequate education (Minorini and Sugarman 1999). Prior to these recent court cases, the focus of most school finance reform had been on resources alone. All states, to one degree or another, use state grants to school districts to partially equalize the fiscal resources districts have available at a given rate of property taxation. In most states, grant formulas distribute aid inversely to the size of each district's per student property tax base but fail to account for differences in costs among school districts, differences that may contribute to varying student performance.

It should be emphasized that providing schools with enough resources to achieve state-imposed student performance goals will not guarantee that schools will actually use those resources effectively to improve student performance. However, once state governments have guaranteed that all school districts have sufficient financial resources to achieve state education goals, then the states can be aggressive in taking steps to intervene in those districts that fail to improve student performance.

To determine the minimum amount of money a school district must spend to achieve a specified improvement in student performance, we estimate an educational cost function using data from elementary and secondary school districts in the state of Texas. Texas is a particularly interesting state to study for two reasons. First, as is now well known, Texas has been administering annual student performance tests to its public school students since 1990 and using these tests as the basis for an accountability system that includes monetary rewards for schools and graduation requirements for individual students (Murnane and Levy 2001). Second, spurred by a series of court challenges to its school financing system, Texas's system of state financial aid to local school districts takes explicit account of many of the factors that studies in other states have found to be systematically related to the costs of education.

The rest of this article is divided into six sections. We start with a brief overview of the school finance system in Texas. Then, in the following section, we derive our cost function and discuss a set of estimation issues. The following section presents the econometric results of our cost function estimation. In the next two sections, we address the question of how state aid formulas could be adjusted to account for differences in costs across school districts. We first discuss the calculation of a cost index that allows us to summarize the results of the estimation and then demonstrate how such an index can be used in a formula designed to guarantee that every district would have sufficient fiscal resources to achieve any state-imposed student performance goals. In the final section, we draw some conclusions.

School Finance in Texas

During the 2001–02 academic year, public schools in Texas educated 4.1 million students. Of the $28 billion of revenues raised for public education in 2001–02, 55 percent came from local sources, 42 percent from the state government, and 3 percent from the federal government (Texas Education Agency 2002). The state of Texas is divided into 1,045 school districts, with 968 providing K–12 education. The state uses a complex mechanism for distributing state aid to these districts. The major elements of the state aid system involve a $250 per student grant to each district; a foundation formula, with the $2,537 per pupil foundation level of spending adjusted for diseconomies of scale and for differences across districts in the cost of resources; and a guaranteed tax base formula for districts with property tax rates in excess of $8.60 for each $1,000 of property valuation and per pupil property tax bases below $258,100. Using a system of pupil weights, additional state aid is provided to school districts with concentrations of children from economically disadvantaged families and children eligible for various special education programs designated for those with disabilities and with limited proficiency in English. Finally, a unique element of the Texas system of school finance is that property-wealthy districts (those with more than $300,000 per weighted pupil) are required to "reduce their wealth to this level." In most cases, this is accomplished by agreeing to educate students residing in other districts or by purchasing "attendance credits."[1]

The Texas Education Authority has argued that student performance on state-administered exams has improved dramatically.

Although there has been considerable debate among scholars concerning the magnitude of improvement in student performance, the Texas Education Authority has argued that student performance on state-administered exams has improved dramatically (Murnane and Levy 2001). Nevertheless, data from testing done during the 2000–01 school year demonstrate that student performance in school districts with a high percentage of poor children and districts with a high percentage of minorities was substantially below average. For example, students in the 87 school districts where more than 75 percent of students came from poor households had composite test scores that were nearly one and a quarter standard deviations below average. Average student performance in some Texas cities was even weaker. For example, the average pass rate in San Antonio was 1.85 standard deviations below average and the average rate in Dallas was two and a quarter standard deviations below average.

[1] See Texas Education Agency (2002) for a full description of these provisions.

Cost Function Estimation

Using data on per pupil school expenditures, student performance, and various characteristics of school districts, we estimate a cost function for K–12 public education in Texas. Estimating a cost function allows us to quantify the relationship between per pupil spending; student performance; various student characteristics; and the economic, educational, and social characteristics of school districts. We follow Bradford, Malt, and Oates (1969) in specifying the output of public schools (measured, for example, by student performance on standardized exams) as a function of school resources, such as teachers and textbooks, the characteristics of the student body, and the family and neighborhood environment in which the students live. This relationship is represented by equation (1), where S_{it} represents an index of school output, X_{it} is a vector of direct school inputs, Z_{it} is a vector of student characteristics, and F_{it} is a vector of family and neighborhood characteristics. The subscript i refers to the school district and subscript t refers to the year.

$$(1) \quad S_{it} = g(X_{it}, Z_{it}, F_{it}).$$

To move from this education production function to a cost function, a relationship between school inputs and educational spending is specified. This is shown in equation (2), where per pupil expenditures, E_{it}, are considered as a function of school inputs, X_{it}, a vector of input prices, P_{it}, and ε_{it}, a vector of unobserved characteristics of the school district.

$$(2) \quad E_{it} = f(X_{it}, P_{it}, \varepsilon_{it}).$$

The final step involves solving equation (1) for X_{it} and then plugging it into equation (2). This gives the cost function represented by equation (3), where u_{it} is a random error term.

$$(3) \quad E_{it} = h(S_{it}, P_{it}, Z_{it}, F_{it}, \varepsilon_{it}, u_{it}).$$

Typically, equation (3) is assumed to be log linear and estimated with district-level data for a given state. In the next section, we present estimates of equation (3) using 1995–1996 data for K–12 school districts in Texas. The dependent variable is the log of per pupil expenditures (excluding spending on transportation). The resulting coefficients indicate the contribution of various district characteristics to the cost of education, holding constant the level of output. There is some discussion in the literature on education production functions about the desirability of using school-level data (Hanushek, Rivkin, and Taylor 1996). We use district-level data here for two reasons: First, state aid in almost all states is distributed to the district, and there is very little systematic information on how money is spent at the school level. Second, several of the school and community variables that we include in our analysis are available only at the district level. In the remainder of this section, we discuss a number of methodological and data issues that must be addressed to carry out the estimation.

This study differs from other cost function studies in the use of a value-added measure of student performance.

As pointed out by Duncombe and Yinger (1999), estimating cost functions provides a practical way to identify and quantify the factors that influence the costs of education, in which the output of school districts can be measured using multiple measures of school performance. Although student performance can, in principle, be measured in various ways, many states measure the effectiveness of schools by relying on standardized exams. For several years, Texas has been testing all students in grades 3 though 8 and in grade 10 in reading and math. The tests are administered in the spring of each year as part of the Texas Assessment of Academic Skills (TAAS). Considerable media attention is paid to the test score results, and improvements in average test scores (or lack thereof) are monitored closely.

One of the ways in which this study differs from other cost function studies is in the use of a value-added measure of student performance in each school district. As a measure of school district output, we compare the average of the composite passing rate on the TAAS exams across grades 4 through 8 and in grade 10 in 1995–1996 with the average passing rates in grades 3 through 7 of the same cohort of students in 1994–1995 and the 8th-grade TAAS passing rate in

1993–1994 (to match the 10th-grade passing rate in 1995–1996).[2] Robert Meyer (1996) provided a strong argument for using a value-added approach to isolate the contribution of school resources to increases in student achievement. He pointed out that the use of average scores from a single grade measures the average level of achievement prior to entering first grade, plus the average effects of school performance and of family, neighborhood, and student characteristics on the growth of student achievement from all years of previous schooling. It is thus likely that rather than providing a measure of the contribution of schools to the growth in student achievement, the single grade score primarily reflects the impact of family and neighborhood environment on student achievement. In addition, many of the recent policy proposals regarding standards, including those of the Bush administration, have focused on improvement in test scores from year to year. The value-added approach is thus more useful for simulating the effects of actual policies.

Educational output variables and per pupil expenditures are determined simultaneously.

In addition to the TAAS scores, we also include student performance on the ACT exams as a measure of the quality of the preparation of students for higher education. Using scores on these exams as a measure of school quality can be problematic, however, because students decide whether to take the exam. Only students with a particular interest in continuing on to college will choose to take these exams, and these are presumably the "best" students, so their scores may reflect their own abilities and motivation rather than any influence of the school. By treating these scores as endogenous, we are able to control for this self-selection. As an instrument for ACT scores, we include the percentage of students who take a college entrance exam.

Estimation of the cost function must take account of the fact that the educational output variables and per pupil expenditures are determined simultaneously. That is, although local school board decisions to raise the level of student performance are expected to have direct implications for the level of spending, decisions concerning per student spending are likely to influence student performance. To deal with this simultaneity, we estimate equation (3) using two-stage least squares, with the school output variables treated as endogenous. As instruments for these school output variables, we draw on a set of variables that are related to the demand for public education. Following a long literature on the determinants of local government spending, we model the demand for public education as a function of school district residents' preferences for education, their incomes, and the tax prices they face for education spending. To the extent that the median voter model provides a reasonable framework for modeling school district spending decisions, it is appropriate to use median income and the tax price faced by the median voter as instruments.[3] We also include as instruments two socioeconomic variables that may be related to the preferences for public education: the percentage of households with children and the percentage of household heads who are homeowners.[4,5]

For school input prices, we focus only on teacher salaries. Teachers are the single most important factor in the production of education, and not surprisingly, teacher salaries account for the largest share of school expenditures. It is important, however, to recognize that teacher payrolls are determined both by factors under the control of local school boards and factors

[2] Test scores represent the same students in the two academic years to the extent that interdistrict student mobility is relatively low. A recent study of elementary school students in Texas by Hanushek, Kain, and Rivkin (2001) found that roughly 86 percent of fourth- to seventh-graders remain in the same district from one year to the next.

[3] We use the tax price implied by Texas's aid formula.

[4] As mentioned previously, we also include the proportion of students who take a college entrance exam as an instrument for ACT scores.

[5] In the results presented in the next section, the 1995–1996 Texas Assessment of Academic Skills (TAAS) scores are treated as endogenous, but the lagged scores are not. Hausman specification tests could not reject the null hypothesis that the lagged scores are exogenous.

that are largely outside of their control. In setting hiring policies, districts make decisions about the quality of teachers they wish to recruit. These decisions have obvious fiscal implications. For example, a district can limit its search for new teachers to those with advanced degrees, those with high grade point averages, or those with a certain number of courses in their teaching specialty. Teacher salary levels are generally determined through a process of negotiation with teacher unions, and school boards have a substantial impact on the outcome of these negotiations. At the same time, the composition of the student body, working conditions within schools, and area cost of living play a potentially large role in determining the salary a school district must offer to attract teachers of any given quality. These factors will be reflected in student and district cost variables, to be described below. We would therefore like a measure of teacher salaries that reflects only salary differences that are outside the control of local school districts. One such measure is the teacher cost index developed by Jay Chambers (1995). Using 1990–1991 data from the National Center for Education Statistics' nationally representative Schools and Staffing Survey, Chambers estimated hedonic wage equations for teachers. He isolated those factors that are outside the control of the local school district (such as the racial composition of the student body, local climate, crime rates, etc.) and used the coefficients for just those factors to construct a teacher salary index for each district in the country. By using this index as our measure of teacher salaries, differences across districts reflect true differences in costs rather than differences in school board choices.

The vectors of student, family, and neighborhood characteristics, Z_{it} and F_{it}, include several variables that influence a district's level of spending per pupil. First, there is considerable evidence that there are higher costs associated with the education of children from low-income families. To measure the number of children from economically disadvantaged families, we use the percentage of students who qualify for the federal government-financed free and reduced-price lunch program or other public assistance. Second, there is a substantial literature that documents the extra costs associated with educating students with various kinds of disabilities and students who enter the schools with limited knowledge of English. Therefore, we include the percentage of students who have been identified as limited English proficient and the percentages of students with two categories of disabilities: the percentage of students who are classified as having any type of disability and the percentage of students who are classified as autistic, deaf, or blind. Third, to reflect the possibility that more resources may be needed to provide a high school education as compared to an elementary school education, we also include the proportion of each school district's student body that is enrolled in high school. Finally, to reflect potential diseconomies of scale associated with both small and large school districts, we include each district's enrollment and enrollment squared. Summary statistics of all variables are presented in table 1.

Some school districts will have higher levels of per pupil expenditures than other districts because those school districts are inefficient.

The variable ε_{it} in equation (3) represents the unobserved factors in each school district that influence district spending. One such factor is the "inefficiency" of the district. That is, even after accounting for differences across school districts in cost factors, input prices, and student performance, some school districts will have higher levels of per pupil expenditures than other districts because those school districts are inefficient. This could mean that they are inefficiently organized or managed or that they use ineffective teaching techniques or employ a particularly ineffective group of teachers.

A number of recent articles have used complex statistical techniques to identify spending that is high relative to spending in districts with similar student performance and costs.[6] Although great care must be taken in interpreting this extra spending as a measure of inefficiency, we include in our cost function estimates an efficiency index calculated using data envelopment

[6] See, for example, Bessent and Bessent (1980); Deller and Rudnicki (1993); Duncombe, Ruggiero, and Yinger (1996); McCarty and Yaisawarng (1993); and Ruggiero (1996).

Table 1. Descriptive statistics for 803 Texas K–12 school districts: 1995–96

Variable	Mean	Standard deviation	Minimum value	Maximum value
Per pupil expenditures, 1995–96, excluding transportation (in dollars)	5,565	1,039	2,907	11,444
Composite TAAS pass rate, 1995–96	79.6	8.3	51.3	96.2
Composite lagged TAAS pass rate, 1993–95	75.5	9.5	37.1	95.4
Average ACT score	19.9	1.6	15	26
Teacher salary index	84.5	9.2	62.2	107.5
Percent of students eligible for free and reduced-price lunch and other public assistance	44.4	18.3	0.1	100.0
Percent of students with disabilities	13.4	4.0	0.7	36.6
Percent of students with severe disabilities	0.21	0.20	0	1.82
Percent of students with limited English proficiency	6.5	10.5	0	72.5
Percent of students enrolled in high school	28.6	3.2	16.2	49.9
Student enrollment	4,081.5	8,614.1	124	74,772
Efficiency index	0.59	0.12	0.3	1
Tax price	0.5	0.3	0	1
Percent of households with children	34.8	7.7	16.2	70
Percent homeowners	73.7	10.3	0.0	99.2
Median income (in dollars)	23,814	7,258	8,196	58,135
Percent of students taking college entrance exams	65.1	14.6	21.4	100

NOTE: TAAS is Texas Assessment of Academic Skills. The teacher salary index is normalized around 1 for all school districts in the United States. As indicated in the table, the average district in Texas has salary costs that are 84.5 percent of the national average. Texas school districts with relatively high teacher costs as measured by the index have costs that are above the national average. The efficiency index takes a value of 1 for the most efficient districts. The data in the table show that the average district in Texas is 59 percent as efficient as the most efficient districts in the state.

SOURCE: Calculated by the authors from data from the Texas Education Agency.

analysis.[7] Data envelopment analysis is a nonparametric estimation procedure that compares each district to a production frontier. Thus, after controlling for student performance and cost differences, lower spending districts are considered to be operating with "best practices," whereas any extra spending may be interpreted as a measure of school district inefficiency.

Cost Function Results

To account for the large variance in district size in Texas, we weight the regressions by district enrollment and drop Dallas and Houston from the sample. Because of missing test scores, we were also forced to exclude 163 of the 968 K–12 school districts. These excluded districts tend to be somewhat smaller, poorer, and higher spending than the 803 districts that remain in our sample and that provide the basis for the cost function estimation.

Recall that we treat the school outcome variables as endogenous and estimate equation (3) using two-stage least squares.[8] The first two columns of table 2 present cost function results that include a measure of school district efficiency, whereas the second two columns are estimated without that variable. The test scores have the expected signs; because lagged scores are a proxy for past levels of students' achievement, high scores mean that districts can spend less to achieve a given level of educational progress. The cost variables generally have the expected signs, and many are statistically significant. Consistent with previous studies, we find a U-shaped relationship between spending per pupil and school district size; with our estimates, the bottom of the U is at roughly 22,026 students when the efficiency measure is included and 9,115 when it is not included. In contrast to the results of some other studies, we find that costs do not appear to be higher for high school students, although that variable is not statistically significant.

7 See Duncombe, Ruggiero, and Yinger (1996) for further discussion of the measurement of school district efficiency using data envelopment analysis.

8 The detailed first-stage regression results can be found in the authors' *Public Finance Review* article.

Table 2. Education cost function for 803 Texas K–12 school districts: 1995–96

	Dependent variable: Log of expenditures per pupil			
Independent variables	**Coefficient**	***t*-statistic**	**Coefficient**	***t*-statistic**
Intercept	3.23**	1.66	–2.29	–0.47
Log of composite TAAS pass rate, 1995–96	3.34*	2.25	7.29*	2.02
Log of lagged composite TAAS pass rate, 1993–95	–2.53*	–2.16	–5.93*	–2.04
Log of average ACT score	1.03*	3.40	1.76*	2.44
Teacher salary index	0.0015**	1.88	0.0031*	2.10
Percent of students eligible for free and reduced-price lunch	0.12	1.64	0.57*	2.92
Percent of students with disabilities	0.02	0.12	0.55	1.42
Percent of students with severe disabilities	3.58	1.03	9.43	1.29
Percent of students with limited English proficiency	0.41*	3.86	0.66*	2.63
Percent of students enrolled in high school	–0.20	–0.63	0.20	0.32
Log of student enrollment	–0.20*	–4.25	–0.31*	–3.24
Square of log of student enrollment	0.01*	3.95	0.017*	3.05
Efficiency index	–1.08*	–9.91	†	†
Sum of squared errors (SSE)	8.571		8.571	

† Not applicable.

* Indicates statistically significant at the 5 percent level.

** Indicates statistically significant at the 10 percent level.

NOTE: TAAS is Texas Assessment of Academic Skills.

SOURCE: Calculations by the authors.

The differences in the cost functions with and without the efficiency measure highlight one of the drawbacks of the technique we use to measure efficiency. Our measure of efficiency captures the effect of all factors that lead spending to be higher than the minimum cost of providing any given mix of public school output. Thus, school districts with above-average spending on things not measured by standardized tests (e.g., advanced music and arts courses) will be characterized as inefficient. Also, higher spending that is attributable to the higher costs of, for example, educating an above-average share of economically disadvantaged students will, in part, be characterized as "inefficiency." As pointed out by Duncombe, Ruggiero, and Yinger (1996), the fact that these higher costs will be attributed in part to the efficiency measure and in part to the cost factors explicitly included in the cost function will mean that the cost function estimates with the efficiency measure will provide an underestimate of the full effects of the cost factors on education spending. This could explain, for example, why the coefficients on many of the cost factors increase when we do not include the efficiency measure. On the other hand, the coefficients in the model without the efficiency measure may be biased upward. We suspect that the "true" cost effects lie somewhere in between those indicated by the cost functions estimated with and without the efficiency adjustment.

In summary, our estimated cost function suggests that in Texas, characteristics of school districts beyond the control of local school officials contribute to the amount of money needed to achieve any given level of student performance. This implies that equal per pupil spending should not be expected to result in equal student performance gains in all districts.

Cost Index Construction

Estimating a cost function provides information about the contributions of various characteristics of school districts to the costs of education. The calculation of a cost index allows for the summarization of all the information about costs into a single number for each district. For example, if we assume that the policymakers in a state define the minimum standard for an accountability system as the current average level of student performance, then a cost index can be constructed that will indicate, for any given district, how much that district must spend, *relative to the district with average costs*, for its students to meet the state's student performance standards.

To demonstrate the calculation of cost indexes, we set the TAAS scores and ACT scores at the average for all Texas districts. As discussed above, we use a value-added measure of student achievement in our cost function. Thus, the coefficient on 1995–96 scores reflects the increase in spending associated with an increase in student performance given an initial level of test score performance in 1994–95. In calculating the cost index, we set the lagged score equal to the average as well; thus, our performance standard is not the average level of student performance but the average gain in performance, that is, the average increase in the percentage of students passing the TAAS exams. The values of the cost factors are allowed to vary for each district, so we predict the level of spending required for each district to achieve this average gain, given their actual costs.

We want to emphasize that alternative performance standards could be used to calculate the cost index; we use the average gain in scores here only as an example. The use of a different standard will not affect the relative ranking of districts in terms of their costs but will change their absolute cost index values and thus will influence any distribution of state aid that depends on the cost index.

Using the cost function estimated without the efficiency index, we calculate that the school district in Texas with average costs (i.e., where each of the cost factors is set equal to its mean) must spend $5,610 per pupil (in 1995–96) to reach our performance standard. For any given school district, the product of this number and its cost index (divided by 100) will indicate the minimum amount that district must spend to meet the student performance goal. Thus, for example, a Texas school district whose cost index is 125 will need to spend $7,012 per student ($5,610 times 1.25) to reach the student performance standard.

The first column of table 3 shows the variation in costs across K–12 school districts in Texas. The district with the lowest costs could achieve average performance by spending about two-thirds as much per pupil as the district with average costs. At the other extreme, the district with the highest costs must spend almost twice as much as the average cost district to provide an average educational outcome for its students. The large range of the index reflects in part the values of the index in a few districts. Ignoring the 10 percent of districts with the lowest index values and the 10 percent of districts with the highest values substantially reduces the range of the cost index. The restricted range in table 3 shows that the district at the 10th percentile has costs that are about 20 percent below average cost and the district at the 90th percentile has costs that are 20 percent above average.

Table 3. Distribution of education cost indexes for 803 Texas K–12 school districts: 1995–96

	Cost index with no efficiency adjustment	Cost index with efficiency adjustment	Texas index
Mean	100.0	100.0	100.0
Median	96.4	98.4	98.0
Standard deviation	17.7	7.1	14.9
Range	124.8	49.8	83.3
Minimum	67.1	86.7	75.1
Maximum	191.9	136.5	158.4
Restricted range	38.0	14.8	36.3
Minimum at 10 percent	82.1	93.0	83.3
Maximum at 90 percent	120.1	107.9	119.6
Correlations:			
Cost index with no efficiency adjustment	1.000	†	†
Cost index with efficiency adjustment	0.959	1.000	†
Texas index	0.513	0.558	1.000

† Not applicable.

SOURCE: Calculations by the authors.

When the estimated cost functions include no measure of efficiency, it is possible that we are interpreting extra spending that is caused by inefficiencies on the part of school districts as higher costs. When a measure of efficiency is included in the calculation of the cost index, the maximum cost index falls from 192 to 136. This suggests that the high cost index numbers for some districts may reflect in part some degree of inefficiency on the part of these local school districts. It is important to emphasize that even after adjusting cost indexes for inefficiency, the variation in costs across districts remains substantial. The correlation between the indexes with and without the efficiency measure is 96 percent, suggesting that including a measure of efficiency has relatively little effect on the rank ordering of districts in terms of costs but can significantly reduce the range. As mentioned previously, however, one must take care in interpreting this difference as entirely attributable to inefficiency.

The school finance system in Texas distributes state aid to local school districts using several formulas that include a number of adjustments for cost differences across districts. Although the formulas do not include a single cost index, they do include separate adjustments for cost-of-living differences, for diseconomies of scale in small and mid-size districts, and for the higher costs necessary to provide education to students from economically disadvantaged families, students with disabilities, and students with limited proficiency in English. Although the cost-of-living adjustments were developed from a careful empirical study, the origin of the other weights and adjustments is unclear. We suspect, however, that the explicit and implicit weights given to each of these cost factors were determined as a result of complex political negotiations and thus are not likely to reflect true cost differences. In contrast, the weighting of each cost factor in our cost index comes from the parameter estimates of the cost function. If our cost function is estimated correctly, these weights indicate the relative contribution of each cost factor to the overall costs of achieving a given student performance standard.

Diseconomies of scale contribute to higher costs in small districts.

To determine whether the current set of cost adjustments used in the distribution of state aid in Texas is compatible with reaching student performance standards throughout the state, we compare our cost index to an *implicit* index generated from the Texas aid program. To construct this index, we add together the basic foundation level (called the "basic allotment" and equal to $2,387 in 1995–96) and the total amount of each district's *special* allotments reflecting all of the cost adjustments mentioned previously (for district size, for student disabilities, etc.). For each district, this sum is converted into an index number by dividing each sum by $3,453, the mean value of these sums across all districts.[9] Summary statistics of the resulting index, labeled Texas index, are shown in the third column in table 3.

Although the range and the restricted range of the Texas index are between the ranges of the two variants of our cost index, the simple correlations between our indexes and the implicit Texas index are relatively low—0.558 and 0.513 for our indexes with and without the efficiency adjustment, respectively. As we shall demonstrate below, there are two important reasons why the indexes differ. First, our index is quite highly correlated with the percentage of children from economically disadvantaged families, whereas the correlation between poverty and the implicit Texas index is much weaker. Second, although our cost function indicates that diseconomies of scale contribute to higher costs in small districts, the aid adjustments for small district size in the Texas aid formulas are much larger than the importance of small size indicated by our cost functions.

The Design of School Finance Formulas

Foundation formulas are currently used by the majority of states to distribute state aid to local school districts. The formulas are designed so that each school district that uses a state-determined "minimum" prop-

[9] In 1995–1996, the cost adjustments and weights used in the state aid formulas resulted in $1,066 in additional state aid (above the basic allotment of $2,387) in the average district. The sum of these two numbers equals $3,453.

erty tax rate will be able to achieve a "foundation" level of per pupil spending. If costs were identical in all school districts, then the state could guarantee that each school district had sufficient resources to achieve the state-specified minimum performance level by defining the foundation level as the spending necessary to produce that particular level of "output."

The results presented in the previous section indicate that costs (at least in Texas) differ substantially among school districts. Thus, to guarantee that all school districts within a state will have sufficient resources to meet state performance standards, it is necessary to develop a foundation formula where the foundation level of spending varies according to differences in costs across districts and where the average foundation level equals the dollar amount necessary to meet the performance standards in districts with average costs.

Costs (at least in Texas) differ substantially among school districts.

A conventional foundation aid formula is presented in equation (4), where A_i equals the foundation aid per pupil in district i, E^* is the foundation level of per pupil spending, t^* the mandated local property tax rate, and V_i the property value per pupil in school district i:

(4) $A_i = \text{MAX}\{E^* - t^*V_i, 0\}.$

To adapt a foundation formula so that it will guarantee that every district has sufficient resources to meet the state's performance standards in our example, measured as the average gain in test scores, requires a determination of the amount of money school districts with average costs need to meet the state standard. Referring to this standard as S^*, $\hat{E}$ can be defined as the amount a school district with average costs must spend to meet the standard. A foundation formula designed to guarantee that every school district has sufficient resources to achieve S^* can be written as

(5) $A_i = \text{MAX}\{\hat{E}c_i - t^*V_i, 0\},$

where c_i is the value of the cost index in school district i.[10] To demonstrate the use of this formula using Texas data, we define $\hat{E}$ as the expenditure needed to achieve the average ACT scores and average TAAS performance gain in a district with average costs. The amount of aid allocated to district i using this cost-adjusted foundation aid formula will be a function of the per pupil property wealth in district i and the relative costs in district i. We have chosen as the foundation level of per pupil spending $5,610, the amount needed to achieve the average performance in a district with average costs. Our choice for the required property tax rate (t^*) is 8.6 mills (or 0.86 percent), which was the actual required mill rate for the first tier of the Texas foundation program in 1995–96.

The El Paso school district can be used to provide an example of the operation of the cost-adjusted foundation formula. In 1995–1996, El Paso had more than 64,000 students, 80 percent of whom were non-White, and two-thirds of whom were from poor families. El Paso's cost index was 29 percent above average when the index was calculated without an efficiency measure and 12 percent above average when the efficiency measure was included. These index values imply that to achieve the average gain in student achievement, El Paso will need to spend between 12 percent and 29 percent more than the district with average costs. State aid could provide these funds by establishing a cost-adjusted foundation level for El Paso between $6,283 (1.12 x $5,610) and $7,237 (1.29 x $5,610).

As discussed in the previous section, various cost factors and pupil weights influence the distribution of state aid in Texas. To focus on how the distribution of state aid would change by replacing the existing weights and adjustments with ones that are derived from our estimated cost function, we conducted several simulations of Texas school aid using a cost-adjusted foundation formula with alternative cost adjustments. The first two columns of table 4 summarize the distribution of cost-adjusted foundation aid using our cost index, without and with the efficiency

[10] See Ladd and Yinger (1994) for a detailed derivation of a cost-adjusted foundation formula.

Table 4. Distribution of aid per pupil for 800 Texas K–12 school districts under alternative cost-adjusted foundation formulas: 1995–96

	Cost index (with no efficiency measure)	Cost index (including efficiency measure)	Cost-adjusted (with Texas index)	Percent difference between average of cost index formulas and Texas index formula
Mean	$4,170	$4,167	$4,154	0.3%
Standard deviation	1,580	1,200	1,325	4.9%
Minimum	0	0	0	0
Maximum	10,560	7,400	7,941	13.1%
Total aid (in billions of dollars)	$13.7	$13.7	$11.9	15.1%
District size quintiles				
1 (smallest)	$4,236	$4,202	$4,385	–3.8%
2	3,737	3,961	3,409	12.9%
3	4,575	4,328	3,725	19.5%
4	3,900	4,038	3,415	16.2%
5 (largest)	4,534	4,280	3,564	23.6%
Percent poor quintiles				
1 (fewest poor)	$2,977	$3,652	$3,431	–3.4%
2	3,557	3,891	3,953	–5.8%
3	3,857	3,943	3,970	–1.7%
4	4,587	4,350	4,475	–0.1%
5 (most poor)	6,407	5,307	5,109	14.6%

SOURCE: Calculations by the authors.

measure. The next column of table 4 shows the distribution of aid using the Texas index. Recall that the Texas index reflects the pupil weights and other cost adjustments used in the actual distribution of state aid in academic year 1995–96. All three simulations use the same foundation level ($\hat{E}$) and required tax rate (t^*).

The simulation results show that including an efficiency measure in the cost function used to construct the index has little effect on the size of the average grant but substantially reduces the magnitude of the largest grant.[11] Distributing grants using the cost-adjusted formulas based on our cost index would have required an aid budget of $13.7 billion. Distributing aid using the Texas index would require an aid budget of only $11.9 billion because the Texas index would distribute the largest per pupil grants to the smallest school districts.

The differences in the pattern of aid distribution can be seen most clearly in the bottom two panels of table 4. In the middle panel, we have divided school districts into pupil-weighted quintiles by district size. Each quintile thus includes approximately the same number of students but a different number of districts. Included in the first quintile are the 595 school districts with enrollment below 3,320, whereas the fifth quintile contains just 11 school districts, all of which have enrollments in excess of 43,550. The data show clearly that the Texas school aid formula allocates more aid to small school districts and considerably less aid to large school districts than would a foundation formula based on cost adjustments derived from our estimated cost functions. Comparing the average aid allocation from our two cost index simulations (with and without the efficiency measure) with the aid allocation from the Texas index simulation, aid would increase by nearly 25 percent in the largest district-size quintile, whereas aid would be reduced by about 4 percent in the smallest size quintile.

The data in the bottommost panel of table 4 divide school districts into pupil-weighted quintiles by the

[11] Ranking per pupil grants by size, the grant at the 90th percentile is $580 larger when the efficiency measure is not included in the cost index calculation compared to the 90th percentile grant when the efficiency measure is included.

percentage of district enrollment that is poor. Although all three simulations generate larger per pupil grants to school districts with concentrations of poor pupils, comparing the aid distributions indicates that our cost functions imply a higher weight on concentrated poverty than the weight given to poverty in the actual Texas school aid formulas. On average, the two cost index simulations generate a 15 percent larger per pupil grant in the highest poverty quintile than the grant generated by the pupil weights used in the existing state aid formulas.

By definition, the first two simulations in table 4 are designed to distribute state aid in such a way that every school district would be provided with an amount of revenue sufficient to enable them to achieve at least the current state average gain in TAAS scores. Our cost function results clearly indicate that improving student performance requires additional resources. It is thus not surprising that the $13.7 billion budget for implementing either one of the cost-adjusted aid foundation formulas will be greater than the amount the state actually spent on school aid. In fact, for the 1995–96 academic year, the state distributed $6.8 billion in state aid to the 800 school districts included in the simulations.[12] This implies that to provide all school districts with sufficient revenues to achieve average gains in TAAS scores would have required a doubling of state aid (actually a 101 percent increase in aid). To put this increase in state aid in context, Texas in 1995–96 was a relatively low-spending state. At $5,473, it ranked 33rd in expenditures per pupil (Snyder 1999). In addition, at 42.9 percent, the state government's share of total education revenue was below the national average. The state share of education funding was higher in 31 other states. If the state government had increased aid to local school districts by 85 percent, the state share would have risen to 60.2 percent of total revenue, a share that would have still been lower than the state share in 11 other states.

There appears to be a consensus that measuring student academic performance is absolutely necessary if the quality of education provided to many of the nation's poor children is to improve.

Conclusions

Policy debates are raging about how student performance should be measured, the type of tests that should be used, and the appropriate role testing should play. Despite strong disagreement concerning the answers to these issues, there appears to be a growing consensus that measuring student academic performance is absolutely necessary if the quality of education provided to many of the nation's poor children is to improve. In this article, we have argued that if states are going to require their students to meet these more rigorous educational goals, they must recognize that achieving these goals will require more resources in some school districts than in other districts for reasons that are outside the control of local school officials. This implies that a necessary, though not sufficient, condition for achieving any given performance goal is that state fiscal assistance to local school districts account explicitly for differences in costs across districts within a state. We have demonstrated that a cost-adjusted foundation formula can be an effective instrument for this purpose.

We use data from Texas to show that it is possible to measure cost differences across districts and that these cost differences are large. We then demonstrate the use of cost-adjusted foundation formulas as a mechanism for distributing state aid in a way that will enhance the chances that a state can meet its student performance goals. In Texas, where cost considerations already play a major role in the distribution of state aid to local school districts, we conclude that reforming the existing state aid formulas to provide a heavier weight to children from economically disadvantaged families and a lower weight to small, mainly rural, districts would better align the distribution of fiscal resources with the underlying costs of education.

It is important to note that the debates over education standards center around two different educational

[12] Because of missing data, three school districts had to be dropped before conducting the aid simulations reported in table 4.

goals. In this article, we have focused on the goal of annual improvement in student performance. But a second goal involves bringing all students (or groups of students, characterized by race, gender, or location) up to a target performance level. Policies in a number of states requiring graduation tests and prohibiting social promotions are examples of absolute student performance standards. The total cost of achieving any absolute standard in a particular school or school district depends in large part on the size of the achievement gap between the current level of student achievement and the standard.

Because the level of student achievement in a number of Texas school districts is substantially below the average level of achievement, it is not surprising that these districts will require a substantial infusion of new resources if they are to close the achievement gap. We recalculated our cost index so that it indicated the cost, relative to the district with average costs, of reaching the statewide average level of student achievement on the TAAS.[13] The range of the resulting indexes increased substantially; the school district with the highest costs had an index value of 718 with no efficiency adjustment, and 220 with the efficiency adjustment. To implement a cost-adjusted foundation formula that would guarantee each school district enough money to reach the average student performance level would require a substantial increase in the size of the state aid budget—to $21.2 billion without the efficiency measure and to $16.9 billion with the efficiency measure.

To guarantee each school district enough money to reach the average student performance level would require a substantial increase in the size of the state aid budget.

In this article, we have demonstrated that the costs of achieving a gains-based standard (in our example, the statewide average annual gain in test scores) will vary substantially across school districts. To ensure that all school districts have adequate resources to sustain annual student performance gains, districts with higher costs will have to be guaranteed additional state fiscal assistance. If annual achievement gains can be maintained, then, over time, low-performing school districts will be able to meet absolute student achievement goals. Obviously, school districts with the lowest levels of current student performance will take the longest time to reach state-imposed standards.

One of the most contentious issues in the debate surrounding the reauthorization of the Elementary and Secondary Education Act was whether low-performing schools should be required to achieve state-imposed performance standards within a fixed number of years. Our estimated cost functions could be interpreted as suggesting that if a school district with high costs is provided with sufficient additional funds, it could fully offset the disadvantages of higher costs in a single year.

We believe that this implication is not justified. In fact, school districts with large achievement gaps will, under most circumstances, take more time to reach any specified state standard than districts with smaller gaps. From a purely statistical point of view, using our estimated cost function to reach conclusions about the money needed to close any given achievement gap within a year generally requires extrapolation beyond the data.[14] More important, in recent years there have been substantial advances in the development of teaching techniques that are effective in improving the academic performance of low-

[13] In the estimation of the cost function, lagged student performance is treated as an endogenous variable because, as with current performance, it is, in part, a choice of the district. In creating the cost index, we want to abstract away from any variation that is under the control of the district. Thus, to account for the endogeneity of the lagged scores, we calculate the cost index using predicted lagged scores, with the predictions based on the coefficient estimates from the first-stage regression, actual values of the cost factors, and state average values for the demand instruments. That is, a district's predicted lagged score reflects the score expected from a district with average preferences and observed cost factors. Put together with the average 1995–1996 score, the level of spending predicted by the cost function is the spending required to reach average achievement given average tastes for education and actual cost factors.

[14] If a high-cost district has a cost index value of 300, this implies that this district will need to spend three times as much per pupil as the district with average costs for its students to reach the student performance standard, say average performance, on the TAAS. Although this conclusion may be correct, assuming that it can be achieved within a single year requires that we use our estimated cost function to extrapolate beyond our data; that is, there are no school districts that achieve average student performance while spending three times the spending level in the district with average costs.

achieving students. Experts on learning recognize, however, that the processing of new knowledge, information, and concepts takes time (Bransford, Brown, and Cocking 2000). Although there is only limited research on the time needed to acquire and process knowledge, it is probably unrealistic to expect that students who are currently performing at substantially below grade level can catch up within a single year even if additional resources are devoted to their education.

Although providing additional financial aid to school districts with large achievement gaps is a crucial step toward reducing those gaps, it is also likely that in many school districts, a large, sudden increase in state aid would not be used effectively to increase student performance (Duncombe and Yinger 2000). Providing new money to schools and school districts with above-average costs, if it is phased in over a period of years, is likely to be most effective in increasing student performance.

Specifying a fixed time limit within which state performance standards must be met and imposing sanctions on those districts failing to meet the deadline is likely to penalize school districts that are currently performing at low levels, even if these districts succeed in making adequate annual progress in improving their students' test scores. Such a policy could lead to discouragement instead of improved achievement. Because an important role of higher student performance standards is to create incentives for schools, teachers, and students to increase the amount of learning that occurs, such standards must be set at reasonable levels.

There are still many issues to resolve in how educational costs and school outputs are measured and in how to reform policy to account for these costs, but it is clear that improving the educational performance of all students requires the annual measurement of student performance, the setting of reasonable goals, and the allocation of state and federal aid to school districts in a way that recognizes differences among school districts both in fiscal capacities and in the costs of providing education.

References

Bessent, A.M., and Bessent, E.W. (1980). Determining the Comparative Efficiency of Schools Through Data Envelopment Analysis. *Educational Administration Quarterly, 16*(2): 57–75.

Bradford, D.F., Malt, R.A., and Oates, W.E. (1969). The Rising Cost of Local Public Services: Some Evidence and Reflections, *National Tax Journal, 22*(2): 185–202.

Bransford, J.D., Brown, A.L., and Cocking, R.R., (Eds.). (2000) *How People Learn: Brain, Mind, Experience, and School*. National Research Council, Washington, DC: National Academy Press.

Chambers, J. (1995). *Public School Teacher Cost Differences Across the United States* (NCES 95–758). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Deller, S.C., and Rudnicki, E. (1993). Production Efficiency in Elementary Education: The Case of Maine Public Schools. *Economics of Education Review, 12*(1): 45–57.

Duncombe, W., Ruggiero, J., and Yinger, J. (1996). Alternative Approaches to Measuring the Cost of Education. In H.F. Ladd, (Ed.), *Holding Schools Accountable: Performance-Based Reform in Education* (pp. 327–356). Washington, DC: The Brookings Institution.

Duncombe, W., and Yinger, J. (1999). Performance Standards and Educational Cost Indexes: You Can't Have One Without the Other. In H.F. Ladd, R. Chalk, and J.S. Hansen (Eds.), *Equity and Adequacy in Education Finance; Issues and Perspectives* (pp. 260–297). Washington, DC: National Academy Press.

Duncombe, W., and Yinger, J. (2000). Financing Higher Student Performance Standards: The Case of New York State. *Economics of Education Review, 19*(4): 363–386.

Hanushek, E.A., Kain, J.F., and Rivkin, S.G. (2001). *Disruption Versus Tiebout Improvement: The Costs and Benefits of Switching Schools.* (NBER Working Paper No. w8479). Cambridge, MA: NBER.

Hanushek, E.A., Rivkin, S.G., and Taylor, L.L. (1996). Aggregation and the Estimated Effects of School Resources. *Review of Economics and Statistics, 78*(4): 611–27.

Ladd, H.F., and Yinger, J. (1994). The Case for Equalizing Aid. *National Tax Journal, 47*(1): 211–224.

McCarty, T.A., and Yaisawarng, S. (1993). Technical Efficiency in New Jersey School Districts. In H.O. Fried, C.A.K. Lovell, and S.S. Schmidt (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications* (pp. 271–281). New York: Oxford University Press.

Meyer, R.H. (1996). Value-Added Indicators of School Performance. In E.A. Hanushek and D.W. Jorgeson (Eds.), *Improving the Performance of America's Schools* (pp. 197–223). Washington, DC: National Academy Press.

Minorini, P.A., and Sugarman, S.D. (1999). Educational Adequacy and the Courts: The Promise and the Problems of Moving to a New Paradigm. In H.F. Ladd, R. Chalk, and J.S. Hansen (Eds.), *Equity and Adequacy in Education Finance: Issues and Perspectives* (pp. 175–208). Washington, DC: National Academy Press.

Murnane, R.J., and Levy, F. (2001). Will Standards-Based Reforms Improve the Education of Students of Color? *National Tax Journal, 54*(2): 401–415.

Ruggiero, J. (1996). Efficiency of Educational Production: An Analysis of New York School Districts, *Review of Economics and Statistics, 78*(3): 499–509.

Snyder, T.D. (1999). *Digest of Education Statistics: 1998* (NCES 1999–036). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Texas Education Agency. (2002). *Snapshot 2002: 2001–02 School District Profiles*. Austin: Division of Performance Reporting, Office of Accountability Reporting and Research. Available at http://www.tea.state.tx.us/perfreport/snapshot/2002/index.html

U.S. Department of Education. (2001). *No Child Left Behind.* Washington, DC: Author. Available at http://www.ed.gov/nclb/overview/intro/presidentplan/page_pg3.html

U.S. Department of Education. (2002). *The No Child Left Behind Act of 2001*: *Executive Summary.* Washington, DC: Author. Available at http://www.ed.gov/nclb/overview/intro/exec-summ.html

Distinguishing Good Schools From Bad in Principle and Practice: A Comparison of Four Methods

Ross Rubenstein

Maxwell School of Citizenship and Public Affairs
Syracuse University

Leanna Stiefel
Amy Ellen Schwartz
Hella Bel Hadj Amor

Robert F. Wagner Graduate School of Public Service
New York University

About the Authors

Ross Rubenstein is associate professor of public administration and senior research associate of the Center for Policy Research in the Maxwell School, Syracuse University. He is also affiliated with the Center's Education Finance and Accountability Project (EFAP). His research focuses on education policy and public finance, specifically, funding equity and adequacy in education, public-sector performance measurement, and merit-based financial aid for college. He was previously on the faculty of the Andrew Young School of Policy Studies at Georgia State University and served as a staff member on Georgia's Education Reform Study Commission. He can be contacted at rrubenst@maxwell.syr.edu.

Leanna Stiefel is professor of economics at the Wagner Graduate School of Public Service, New York University. Her areas of expertise are school finance and education policy, public and not-for-profit financial management, applied economics, and applied statistics. Some of her current and recent research projects include costs of small high schools in New York City; measurement of efficiency and productivity in public schools; segregation, resource use, and achievement of immigrant schoolchildren; racial impact of school finance funding; test score gaps among schoolchildren; and patterns of resource allocation in large city schools. Professor Stiefel is the author of *Statistical Analysis for Public and Non-Profit Managers* (1990) and co-author of *The Measurement of Equity in School Finance* (1984). Her work also appears in journals and edited books. She is past president of the American Education Finance Association, a governor of the New York State Education Finance Research Consortium, and a member of the National Center for Education Statistics Technical Planning Panel (U.S. Department of Education). She can be contacted at leanna.stiefel@nyu.edu.

Amy Ellen Schwartz is professor of public policy at New York University's Wagner Graduate School of Public Service. Her research interests span a wide range in K–12 and higher education. Since earning a Ph.D. in economics from Columbia University in 1989, she has conducted research on education spending, school aid, private support for public schools, and the evaluation of school reforms. Her current research in K–12

education, joint with Leanna Stiefel, focuses on the measurement of school performance, the education of immigrant and native-born children in urban schools, and the Black-White test score gap. This work is supported by the U.S. Department of Education, the Spencer Foundation, and the New York State Education Department. Her work has been published in a variety of journals including *The Journal of Education Finance, American Economic Review, Public Administration Review, National Tax Journal,* and *Brookings-Wharton Papers on Urban Affairs.* She can be contacted at amy.schwartz@nyu.edu.

Hella Bel Hadj Amor is a doctoral student and adjunct professor at New York University's Wagner Graduate School of Public Service. She specializes in education policy, public finance, and applied econometrics. She holds an M.A. in economics from New York University and a B.S. in applied economics from the University of Paris IX Dauphine. She has worked on various projects dealing with issues of education policy and education finance. In particular, she has worked for the Campaign for Fiscal Equity and interned at the regional Mission of the World Bank in Bangkok, Thailand, and the National Bank of Cambodia. She is currently part of a team studying the measurement of efficiency in public schools in New York City and Ohio. Her dissertation work deals with the cost of education in Ohio, and the impact of the size and structure of Ohio school districts on school performance. She can be contacted at tennessee_day@yahoo.com.

Distinguishing Good Schools From Bad in Principle and Practice: A Comparison of Four Methods

Ross Rubenstein

Maxwell School of Citizenship and Public Affairs
Syracuse University

Leanna Stiefel

Amy Ellen Schwartz

Hella Bel Hadj Amor

Robert F. Wagner Graduate School of Public Service
New York University

Acknowledgments

The authors wish to thank Duc-Le To and William Fowler, Jr., of the U.S. Department of Education for helpful suggestions and comments. Sonali Ballal provided expert research assistance. The U.S. Department of Education, Institute of Education Sciences, generously provided support for this research through Field Initiated Studies Grant #R305T010115. The views expressed in this paper do not necessarily reflect those of the Department of Education. All responsibility for omissions and errors is ours.

Introduction

For over a decade, perhaps no other issue in education has generated the same level of debate and policy activity as school accountability. At their most basic, accountability policies tie school rewards and sanctions to measures of school performance, typically specified as either performance levels (for example, aggregate percentile ranks or the percentage of students meeting specified benchmarks) or changes in performance (for example, increases in aggregate test scores or in the percentage of students meeting benchmarks). While most accountability efforts have been enacted at the state and local level, the peak of this movement may be the federal No Child Left Behind (NCLB) Act of 2001, which requires states to demonstrate adequate yearly progress in reading and mathematics performance by school and by subgroups within schools. Common to these reform efforts is the underlying notion that incentives based upon measures of school performance will spur improvements in student performance.

Given the popularity of accountability reforms around the country, it is not surprising that considerable attention is being paid to evaluating the impact of these reforms, both intended and unintended, and assessing the incentives and disincentives embedded in those reforms. (See, for example, Cullen and Reback 2002; Figlio and Winicki 2002; Figlio 2003). In contrast, relatively little attention has been paid to identifying and specifying valid and reliable measures of school performance, even though performance measurement lies at the heart of these reforms. Developing appropriate methods is clearly necessary—though not sufficient—to creating and implementing accountability systems that function as policymakers intend. This paper examines alternative methods of measuring school performance, considering both practical and conceptual issues and evaluating the relative merits of these different measures in applications using data on schools in New York City and Ohio.

Although largely overlooked in the implementation of most accountability reforms, one of the most difficult challenges to be overcome is the difficulty of comparing schools that educate diverse and differing student populations and work with different levels of resources and institutional constraints. Put simply, it has been well established since at least the time of the Coleman report (1966) that variations in student performance, particularly as measured by standardized test scores, are highly correlated with students' socioeconomic backgrounds. On average, students from higher socioeconomic backgrounds do better. Thus, performance measures that fail to account for these student differences risk rewarding schools whose task is in many ways "easier" because of the out-of-school advantages that students from wealthier households tend to enjoy, and because of the potential peer effects generated when the distribution of these students is clustered in certain schools. Likewise, schools serving primarily students without these advantages may appear to be low performing due largely to factors outside their control. Similarly, schools in many urban and rural areas may be labeled low performing in part because they lack the necessary resources to meet the performance levels of schools in wealthier areas.

Variations in student performance are highly correlated with students' socioeconomic backgrounds.

Of course, while most accountability systems focus on the level of performance (say, on standardized tests), school efficiency (in using resources to produce desired outcomes) is perhaps even more crucial in today's constrained fiscal environment. While the sets of high-performing and highly efficient schools will often overlap, schools with generous resources may be able to achieve high performance without optimizing the use of their resources. As state and local budgets bind tightly enough that shortening the school week is seriously considered (Reid 2002), finding schools that make the most effective use of their limited resources (and learning how they do so) becomes increasingly important.

In this paper, we describe, analyze, and compare four alternative methods of measuring school performance and efficiency. Using data from schools in one state (Ohio) and one large city (New York City), we use these different methods to estimate the relative efficiency of public schools and to explore the similarities and differences between the results obtained. To be specific, we explore the extent to which the resulting efficiency measures differ from one another, particularly in the identification of "good" and "bad" schools. Further, we analyze how and why the methods differ, both in their theoretical underpinnings and in their practical applications, and highlight strengths and weaknesses of each method. The remainder of this paper proceeds as follows. The next section describes previous work on the measurement of school performance and efficiency, including conceptual and empirical issues raised by the research. This is followed by brief overviews of the data and the four techniques employed in this paper: adjusted performance measures (APMs), data envelopment analysis (DEA), educational production functions (EPFs), and cost functions. A final section presents the results of analyses using the four methods with similar data sets, compares the results, and presents conclusions on the use of these methods for school performance and efficiency measurement.

Background and Literature

In recent years, a relatively small body of research has begun to accumulate that considers conceptual and empirical issues raised by efforts to measure school performance, specifically in the design and implementation of accountability systems. For example, Hanushek and Raymond (2002) and Ladd (2002) review a number of questions that such policies raise: Do the performance measures reflect the material taught? Is the performance of all students, teachers, and administrators taken into account? What are the most appropriate target scores or rates of increase? What incentives and disincentives are embedded in the system? What data are needed and how do errors in the data affect the performance assessment? While many researchers acknowledge that comparing schools based on average test scores may be unfair because schools may be held responsible for factors (like student background) over which they have no control, deciding which factors should be controlled is also a matter of considerable controversy. For example, there may be general agreement that schools serving high proportions of students from low-income families face greater challenges in generating a high level of student achievement, but there is less agreement that it is appropriate to control for race (Clotfelter and Ladd 1996; Ladd 2002; Ladd and Walsh 2002). Problems may arise when using changes in test scores as output measures because two cohorts of students within the same school may differ in background and cumulative school inputs, thereby producing biased measures (Hanushek and Raymond 2002; Linn and Haug 2002; Ladd and Walsh 2002). Mobility, exemptions from testing, and measurement error may also lead to poor measures (see Clotfelter and Ladd [1996], Hanushek and Raymond [2002], Ladd [2002], and Ladd and Walsh [2002] on mobility; Hanushek and Raymond [2002], and Kane and Staiger [2002], on exemptions from testing; and Hanushek and Raymond [2002], Ladd [2002], Ladd and Walsh [2002], and Linn and Haug [2002] on measurement error).

Schools that experience a gain between two years may not have better educational practices than others do.

Clotfelter and Ladd (1996) observe performance incentive programs in Dallas and South Carolina to determine whether they can and do affect performance. The authors look at various measures of school performance, some based on changes, others on residuals. Overall, they find that measures based on changes in scores are highly correlated and those based on residuals are correlated, but correlations between changes and residuals are much lower. Ladd and Walsh (2002) assess the value-added approaches used to measure school effectiveness in South Carolina and North Carolina. Both states employ simple value-added measures based upon data on current and past student test scores, but do not include adjustments for family background and school-level resources. Using data for 1993–95 on the reading and mathematics fifth-grade test scores of more than 37,000 North Carolina students, they investigate the impact of specification and measurement errors on school rankings and find that the ranking of schools is somewhat sensitive to specification—rankings derived from fixed effects regressions differ from those based upon the mean of residuals over the years of the data. More importantly, using an instrumental variable approach, they find that measurement error bias is responsible for about two-fifths of the higher performance of schools with more advantaged populations and correcting for measurement error causes dramatic changes in the relative rankings of schools.

Kane and Staiger (2002) describe the statistical properties of a variety of performance measures using the performance of fourth-graders taking a mathematics test in 1,163 elementary schools in North Carolina in 1998. They also look at changes between 1998 and 1999 and between 1994 and 1999. The authors argue that accountability systems are often based on imprecisely measured test scores, usually for one year, and that using (weighted) averages of test scores over a few years would be more reliable. By measuring the correlation between changes from one year to the next and seeing whether changes one year are reversed the following year, the authors estimate that at least three-quarters of the variance in test scores is transitory and that small schools, in particular, are more apt to witness such transitory changes. The authors suggest grouping schools by size and distributing rewards/sanctions within each group or giving smaller awards to more schools.

Linn and Haug (2002) confirm Kane and Staiger's findings. Using fourth-grade reading data on Colorado schools for 1997–2000, the authors find that schools with high percentages of high-achieving students have smaller gains than other schools. They also find that schools that experience a gain (or loss) between the first two years generally observe a loss (or gain) between the next two years. Thus, schools that experience a gain between two years may not have better educational practices than others do. And if a school experiences a loss and receives assistance, this assistance may not be responsible for the gain between the next two years. The authors suggest including information on reliability in accountability reports.

Overview of Four Methods

Our work builds upon previous research to examine the properties of four quantitative techniques that may be used to measure school performance: adjusted performance measures, education production functions, data envelopment analysis, and cost functions. Each of these methods allows some way of accounting for differences in the inputs to the educational process across schools—primarily the characteristics of students and resources available to school personnel. Each attempts to measure school efficiency; that is, the school's contribution to producing outputs with a given mix of students and resources, and each relies upon test scores to measure output. The methods differ in their theoretical underpinnings, their data requirements, their ability to include multiple outputs simultaneously, and in the information they provide regarding potential sources of inefficiency. A unique contribution of this paper is that we estimate efficiency scores using each of the methods based upon the same data, allowing us to isolate the differences between the methods (and the information they provide) from differences between data sets and variable definitions. In addition, we use real rather than simulated data, which permit us to explore the many practical issues that arise in using administrative data for these purposes. The next sections review the data sets we use and then briefly discuss each of the quantitative techniques.

A unique contribution of this paper is that we estimate efficiency scores using each of the methods based upon the same data.

Data

For our New York City analyses, we employ a rich school-level database that includes information on student characteristics, test scores, and school resources. The DEA and APM analyses use data for the 1999–2000 school year only, while the EPFs and cost function analyses use data on a balanced panel of 602 elementary schools for 1995–96 through 2000–2001. (A balanced panel includes multiple years of data on the same schools.) The panel includes only schools with third, fourth, and fifth grades and valid reading and mathematics scores for each grade in each year. In addition to school-level aggregates, grade-level demographic variables (race/ethnicity, immigrant status, free and reduced-price lunch eligibility) were calculated from student-level data.[1]

All test score data are reported as standardized *z*-scores. Data for the third and fifth grades come from the CTB/McGraw Hill Test of Basic Skills (CTB) in reading and the California Achievement Test (CAT) in mathematics, while fourth-grade data for 1998–99 and 1999–2000 are from New York State English Language Arts (ELA) reading and mathematics tests. For comparability, the tests are normalized to New York City–wide averages.[2]

In Ohio, the DEA and APMs use data for the 1997–98 school year while the EPFs and cost functions use a panel of 783 schools that include both fourth and sixth grades during a 4-year period, 1995–96 through 1998–99. Passage rates on fourth- and sixth-grade writing and mathematics proficiency exams were used to capture outputs.

Table 1 displays descriptive statistics for both the Ohio and New York City school samples. In developing the databases, we made every effort to use identical sets of schools and variables for each method. Our data sets and variables lists are very similar for each method, though not identical for a variety of reasons, which are discussed more fully in the "Data Challenges" section below.

Adjusted Performance Measures (APMs)

APMs use a regression-based technique in which an output, typically some type of test score measure, is regressed on a set of variables thought to represent factors outside the control of the school itself. These exogenous factors often include student and school characteristics, typically measured as school-level—or perhaps grade-level—aggregates. The APM is, then, each school's estimated residual value, or the differ-

[1] For greater detail on the data, see Schwartz and Zabel (2003) and Schwartz, Stiefel, and Bel Hadj Amor (2003).

[2] Greater detail on the normalizing procedure is available in Stiefel, Schwartz, Bel Hadj Amor, and Kim (2003).

Table 1. Descriptive statistics for New York City and Ohio schools: 1997–2000

	Mean	Standard deviation
New York City schools, 1999–2000		
Z-score, mean fifth-grade mathematics	0.04	0.47
Z-score, mean fifth-grade reading	0.03	0.42
Lagged *z*-score, mean fourth-grade mathematics	.03	0.50
Lagged *z*-score, mean fourth-grade reading	.03	0.46
Grade 5, percent free lunch eligible	73.23	24.94
Grade 5, percent reduced-price lunch eligible	6.86	5.52
Grade 5, percent Black	36.32	32.56
Grade 5, percent Hispanic	33.94	26.05
Grade 5, percent Asian	11.39	16.32
Grade 5, percent with Language Assessment Battery (LAB) score less than 40th percentile	7.58	7.59
Teacher-pupil ratio	0.08	0.01
Non–classroom teacher expenditures (in dollars)	5,653	1,376
Percent of teachers with greater than 2 years' experience in same school	64.11	13.87
Percent of teachers with master's degree	77.27	13.64
Percent of teachers with greater than 5 years' experience	57.66	13.44
Percent of teachers permanently licensed or assigned	81.70	15.00
Enrollment	802.86	339.70
Ohio schools, 1997–98		
Percent passing, sixth-grade mathematics	48.48	20.88
Percent passing, sixth-grade reading	52.60	18.40
2-year lagged percent passing, fourth-grade mathematics	43.38	20.54
2-year lagged percent passing, fourth-grade reading	44.10	18.19
Percent free or reduced-price lunch eligible	28.33	24.44
Percent Black	13.41	25.96
Percent Asian	2.15	5.04
Instructional expenditures per pupil (in dollars)	3,465	606
Noninstructional expenditures per pupil (in dollars)	1,795	490
Enrollment	426.58	178.74

SOURCE: Authors' calculations based upon data provided by the New York City Board of Education, the Ohio Department of Education, and Ohio Education Association.

ence between the actual school output and the output predicted from the regression equation (see Stiefel, Schwartz, Bel Hadj Amor, and Kim [2003]; Rubenstein, Schwartz, and Stiefel [2003]; and Stiefel, Rubenstein, and Schwartz [1999] for more on APMs).

Prior-year test scores may be included as independent variables in order to approximate the school's "value added" to student achievement over the course of the school year. An alternative approach is to measure the dependent variable as the change in test score between years. Resources within the control of the school may also be included in the equation to minimize bias from omitted variables. If such variables are included, they should be set to the sample mean rather than the observed value for each school when calculating the APM. This approach can be used to predict performance given observed factors outside the control of the school and average controllable resources (see Rubenstein, Schwartz, and Stiefel 2003).

While the APM procedure is the most straightforward and "user-friendly" of the techniques discussed here, it is important to note that APMs implicitly assume that all of the estimated error reflects relative school efficiency or inefficiency. To the extent that the error term captures other factors, such as measurement error or the effects of unobserved or omitted variables, the residuals may under- or overestimate school efficiency. Another potential drawback is that, like most regression-based techniques, APMs can be calculated for only one output measure at a time. This does not, however, preclude the analyst from creating a composite measure combining multiple APMs, perhaps standardizing the residuals if the measurement scales differ.

Data Envelopment Analysis (DEA)

DEA is a non-stochastic technique for assessing relative technical efficiency across organizations.[3] More specifically, it requires the construction of a nonparametric efficiency frontier based on the observed input/output ratios of units in the sample, such that all of the efficient decisionmaking units (DMUs) lie on the frontier and "envelop" the inefficient units lying off the frontier. The efficient units receive an efficiency score of 1 (or 100) while each inefficient unit's efficiency is calculated as one minus its distance from the efficiency frontier. Thus, lower scores indicate lower levels of efficiency. The DEA concept differs from regression-based techniques in several ways. First, DEA assesses efficiency in relation to the *best* results actually achieved by units in the sample, rather than the *average* results achieved. Second, DEA can include both multiple inputs *and* multiple outputs. Finally, the DEA procedure seeks to maximize each unit's efficiency rating by assigning unit-specific weights in the linear program. Therefore, units can achieve efficiency through specialization as well as through high performance across multiple measures.

DEA is not a statistical technique and it does not produce coefficients that can be used for testing significance or inferring from this sample to populations. Unlike a more standard production function, though, DEA does not assume that the functional form is the same across schools. The frontier is constructed from the observed inputs and outputs of the units in the sample (Charnes et. al 1994).

A clear advantage of DEA for assessing school efficiency is that it can explicitly account for schools' multiple outputs. However, since schools can reach the frontier by specializing in certain areas, some schools may be deemed efficient despite low performance in some areas. The DEA technique also permits inputs to be labeled as "controllable" or "uncontrollable" to school personnel. And while regression-based techniques may provide little guidance about ways to improve efficiency, DEA produces "slack" values suggesting the reduction in inputs that would be possible without harming outputs.

Education Production Functions (EPFs)

Education production functions link outputs and inputs in a relationship that can be written to include a specific term to capture the persistent efficiency of a school's production process over a period of years. Given panel data on a set of schools for a number of years, efficiency is directly measured and not inferred from the error term, as in the APM. Instead, a "fixed effects model" can be estimated using Ordinary Least Squares (OLS) regression in which the "school effect" captures the impact of the time-invariant characteristics of the school on the output measure, conditional on the observed differences in school inputs and student characteristics. The output may be specified as the change in test scores across grades between years, the change within the same grade between years, or the level of the test score with a lagged-year test score included as a right-hand-side "input." The model may also include a grade effect to capture grade-specific phenomena as well as nonlinear relationships. This formulation is further developed in Schwartz and Zabel (2003).

As with the APM, EPFs permit only one output measure in each equation. Multiple grades or subject areas could, however, be combined in a variety of ways, either as a single composite output measure or by combining school fixed effects from multiple equations. In the EPF, the schools with the largest estimated school effects would be considered the most efficient or "best" schools. Unlike an APM, the fixed effects specification allows the analyst to disentangle the effect of unchanging school characteristics from random error. However, the school fixed effect may still largely be a "black box," capturing all the unmeasured school characteristics affecting performance but offering little guidance as to what those characteristics might be. Another alternative is to "purge" the estimated fixed effects of time-invariant

The schools with the largest estimated school effects would be considered the most efficient or "best" schools.

[3] Extensions of the model can also account for allocative efficiency, but the basic model does not.

characteristics, such as location, by running a second regression. This is explored in Schwartz and Zabel (2003).

Cost Functions

School cost functions are the analogs of production functions, and rest upon the same underlying theoretical foundation. While a production function captures the relationship between inputs and outputs directly, the corresponding cost function captures the minimum cost of producing a given level of output, conditional on the prices of inputs. In principle, the dependent variable in a cost function is the cost incurred in producing education, and the independent variables are outputs, input prices, and other cost factors. In practice, expenditures are used as the dependent variable in cost functions and, as described below, data on input prices are limited.

Cost functions offer several conceptual advantages over EPFs for measuring school efficiency. First, while input quantities may not be exogenous to schools, input prices are more likely to be exogenous at the school level. This is important to interpreting the results—OLS regressions are correctly specified and have the usual interpretation only if the independent variables are exogenous. Otherwise, coefficients can be biased, confounding interpretation. Second, like DEA, cost functions can include multiple outputs simultaneously since they are independent variables in the model. To the extent that input prices are influenced by school activities, though, treating them as exogenous variables may not be appropriate.[4] From a practical standpoint, a more difficult problem is that good data on input prices may be difficult to obtain, particularly at the school level.

The cost function captures the minimum cost of producing a given level of output.

Empirical Analyses and Comparisons of Results

In this section we assess the reliability and consistency of school performance measures across quantitative techniques.[5] As described above, we constructed data sets for use in each analysis with an eye toward consistency across analyses. While it was not possible to construct *identical* data sets because of differing data needs, each analysis uses largely the same schools and same variables as inputs and outputs. We begin with a discussion of issues confronted in assembling the data required for estimating efficiency using the different methods.

Data Challenges

Each of the four methods requires somewhat different data—in variables and in the amount of data (number of years)—and imposes somewhat different limitations on the use of data, implying that slightly different samples will be required. To begin, the methods differ in the treatment of missing data. A basic requirement of all of these methods is that, while some are able to accommodate missing values in independent variables, in all cases observations with missing data for the dependent variable must be omitted. Thus, samples may differ because of differences in the dependent variable used. APM- or EPF-based measures require complete data on the test score specified as the dependent variable; cost-function-based measures require complete data on costs. Rather than restrict our analyses only to those schools for which a full set of data was available with no missing values, which may be unrepresentative of the whole, we allowed for slightly different samples.[6]

4 If, for example, low-performing schools are authorized to pay teachers higher salaries, then the salaries are not appropriately viewed as exogenous, complicating the estimation of cost functions.

5 Papers by Stiefel et al. (2003), Schwartz and Zabel (2003), Rubenstein (2003), and Schwartz, Stiefel, and Bel Hadj Amor (2003) discuss specific issues raised by the analyses using APMs, EPFs, DEA, and cost functions, respectively, and some of these results are also summarized in the conclusions to this paper.

6 For APMs and production and cost functions, missing values in independent variables can be dealt with by interpolation or, as in the studies underlying this paper, by "re-coding" the variables: a new variable is used in the regressions which equals the original variable, if it is not missing, or zero if it is missing. In addition, a dummy variable coded one if the value is reported and zero otherwise is included in the models. The coefficient on this variable, if significant, indicates that the value of the dependent variable varies systematically between the group of schools that have the data and the group of schools that do not.

In this light, APMs and production functions on the one hand complement cost functions on the other: since test scores are *independent* variables in a cost function, schools with missing data on one or more test scores remain in the sample and efficiency measures can be computed for them. Likewise, it is possible to construct efficiency measures for schools that do not have expenditure data by using APMs or production functions.

Dealing with missing data in DEA is more problematic. Observations with missing data cannot be included in DEA, however, and schools that have missing data on any of the variables used are not used for estimation. This again raises an issue of internal validity. In addition, DEA cannot accommodate variables with zero or negative values. Instead, variables may be recoded: a zero is replaced with a very small number (to be defined on a case-by-case basis). Similarly, the DEA procedure assumes that increases in inputs lead to increases in outputs. Thus, if an input is negatively correlated with an output, it must be respecified to have a positive correlation. For example, rather than using the percentage of students who are receiving free lunch as a measure of poverty, the percentage of students who are *not* receiving free lunch is created and included. The latter implies making assumptions as to the relationship between the input and the output. While it is generally accepted that the relationship between free lunch eligibility and school performance is negative, other relationships, say between school performance and the performance of females, are less obvious and must be determined empirically. In practice, assumptions are made based on theory and correlations among the relevant variables. Care must be taken when interpreting the results, however, because a variable can be coded differently in different samples or when using different techniques (such that the percentage of students who are female may be used in one sample and the percentage of students who are male used in another).

The methods differ in the number of years of data required for estimating efficiency.

As noted earlier, the methods differ in the number of years of data required for estimating efficiency. APMs and DEA are cross-sectional methods, which imply that only one year of data is necessary or must be chosen for estimation. While the most current year of data may be used, in an effort to better reflect current conditions, the choice may be based upon the availability of data, if, for example, some variables are available on an irregular basis. If the goal is to compare across methods and/or locations, the most current year of data for *all* methods and/or places may be appropriate. Efficiency measures can be estimated using these methods for several years for comparison across years.

While production and cost functions may be estimated with a single year of data, estimating efficiency measures using school fixed effects, as described above, requires a panel data set. The implication is important: data must be available for a school for at least two years (and more specifically, they must have data on the dependent variable for at least two years), so that new schools must be excluded. In truth, this may be consistent with policy objectives, for example, to give new schools one or two "experimental" years before they are held accountable for student performance. But it may also blunt the incentives for schools to be efficient if the schools are not eligible for rewards or sanctions. More generally, these methods may be somewhat less useful for jurisdictions in which there are a considerable number of school reorganizations, openings, closings, etc., and analysts must make hard decisions about when a school is "new" and when a school is more appropriately treated as persisting, even if it is somewhat changed.

Notice, also, that the methods differ in the variables necessary for estimation—and, while some of these variables are relatively common, others are quite scarce. While the minimum data requirements to use the APM method are relatively easy to meet, the data required to estimate cost functions are rarely available. Even where school-level data on expenditures are available (and ignoring the potential distinction between cost and expenditure data), data on input prices are scarce. Data on teacher salaries, or salary schedules, is crucial for estimating cost functions, yet these are frequently unavailable or, as in our New York City sample, salaries may not vary across the sample of schools, making it impossible to include them in a regression equation.[7] One alternative, perhaps not fully satisfying, is to include teacher characteristics as proxies as we do in our New York City analyses below. In this case, however, the resulting efficiency measures are closer to "adjusted cost measures"—bearing the same relationship to efficiency measures based on cost functions as APMs bear to efficiency measures based on EPFs. In what may be viewed as a "best case scenario," salary data may be available, as in our Ohio sample, but since teacher contracts are typically negotiated at the district level rather than at the school level, salaries are unlikely to vary across schools within districts. Thus, estimating a true cost function requires, at the very least, data that span a significant number of school districts.

Finally, estimating efficiency measures using any of these techniques requires defining a sample of schools over which variables and the estimated parameters are likely to be consistent. Is it plausible that the "technology" of producing education is the same in elementary schools and high schools? If not, then it is inappropriate to estimate the efficiency of these schools as a group—the coefficients of the education production function would be mis-estimated, as would the parameters of the cost functions. While it is clearly difficult to justify comparing, say, elementary and high schools, more subtle choices must be made. As an example, is it appropriate to compare the efficiency of all the schools that serve a sixth grade—using a single production function, APM equation, or cost function—even though some are elementary schools serving kindergarten and other early childhood grades, while others are middle schools serving higher grades? Should we only compare schools having the same grade spans? Doing so would further restrict sample size and make it difficult to form samples of sufficient size. Ideally, a compromise can be found in which all schools in the sample are similar along a number of lines and the sample size is large enough to obtain reliable results—with other differences controlled through the regressions.

Empirical Comparisons

Table 2 displays the Pearson correlation coefficients across the four methods for the Ohio data. Note that two of the methods—APMs and EPFs—have separate reading and mathematics results because, unlike DEA and cost functions, they use only one output measure at a time. The raw correlations among methods, even as different as DEA and APMs or EPFs, are often above 0.5. In fact, only the cost functions exhibit correlations so low with any other method as to be indistinguishable from zero. *Raw* measures for the same grade and same test over years, or different tests in the same year, are often correlated above 0.90. Still, these correlations in table 2, across different methods, are quite high.

[7] This is a minimum requirement. Ideally, data on other input prices should be included, but those are not commonly reported.

Table 2. Ohio Pearson correlation coefficients across four quantitative techniques: 1995–99

	EPF (reading)	EPF (mathematics)	DEA	APM (reading)	APM (mathematics)	Cost function
EPF (reading)	1.000	0.573***	0.509***	0.628***	0.385***	–0.045
EPF (mathematics)	0.573***	1.000	0.505***	0.381***	0.679***	–0.026
DEA	0.509***	0.505***	1.000	0.639***	0.639***	–0.051
APM (reading)	0.628***	0.381***	0.639***	1.000	0.498***	–0.039
APM (mathematics)	0.385***	0.679***	0.639***	0.498***	1.000	0.006
Cost function	–0.045	–0.026	–0.051	–0.039	0.006	1.000

*** Indicates significance at the 1 percent level.

NOTE: Data for DEA and APMs are from 1997–98; EPFs and cost functions use a balanced panel, 1995–96 through 1998–99.

SOURCE: Authors' calculations based upon data provided by the New York City Board of Education, the Ohio Department of Education, and Ohio Education Association.

The results for New York City, displayed in table 3, are somewhat more variable than those for Ohio. While some correlations are higher than those in Ohio (between the reading and mathematics EPFs or between DEA and APMs), many are lower. Once again, the cost function efficiency measures show zero correlation with any other measures.

These correlations may raise more questions than they answer. For example, do correlations in these ranges mean that if a jurisdiction tried to categorize schools into several groups—one indicating successful schools, another indicating failing schools, and a third in the middle—the use of alternative methods would lead to large differences in the schools in the successful and failing groups? As another example, what are the characteristics of the schools that shift groups across methods? And finally, what are the key differences among these methods that should lead a jurisdiction to choose one or the other depending on its objectives? Tables 4 and 5 summarize the results from an array of cross-tabulations based on school rankings from selected methods.[8] The schools are divided into the top group (schools whose efficiency is above the 90th percentile), the bottom group (schools whose efficiency is below the 10th percentile) and the middle group (schools whose efficiency ranges from the 10th to the 90th percentiles). In each cell, a series of numbers indicates the percentage of the schools that move, or do not move, from one of the percentile groups to another when the method listed in the column, rather the method listed in the row, is used.

More specifically, D2 ("down two") is the percentage of schools that move from the top to the bottom ranking group. The top left cell in table 4, for example, indicates that only 0.1 percent of the schools that are at the top based on the reading EPF are at the bottom based on the mathematics EPF. The D group ("down") combines two groups of schools: (1) the schools that move from the top to the middle group and (2) the schools that move from the middle to the bottom group. The same cell indicates that 11.4 percent of the schools move from top to middle or middle to bottom when switching from the reading EPF to the mathematics EPF. C ("constant") is the percentage of schools that are ranked in the same percentile group according to both methods, and is, therefore, the combination of three groups: (1) the schools that are in the top according to both methods, (2) the schools that are in the middle according to both methods, and (3) the schools that are at the bottom according to both methods. The top left cell in table 4 indicates that the vast majority of schools (76.8 percent) are ranked in the same percentile group by the reading and mathematics EPFs. The U group ("up") combines two groups of schools: (1) the schools that move from bottom to middle and (2) the schools that move from middle to top. As shown in the top left cell of table 4, 11.7 percent of the schools

[8] A complete set of cross-tabulations is available from the authors.

Table 3. New York City Pearson correlation coefficients across four quantitative techniques: 1995–2001

	EPF (reading)	EPF (mathematics)	DEA	APM (reading)	APM (mathematics)	Cost function
EPF (reading)	1.000	0.888***	0.168***	0.374***	0.310***	0.037
EPF (mathematics)	0.888***	1.0000	0.157***	0.331***	0.456***	0.086**
DEA	0.168***	0.157***	1.000	0.073*	0.087**	–0.061
APM (reading)	0.374***	0.331***	0.073*	1.000	0.585***	–0.035
APM (mathematics)	0.310***	0.456***	0.087**	0.585***	1.000	0.062
Cost function	0.037	0.086**	–0.061	–0.035	0.062	1.000

* Indicates significance at the 10 percent level.

** Indicates significance at the 5 percent level.

*** Indicates significance at the 1 percent level.

NOTE: Data for DEA and APMs are from 1999–2000; EPFs and cost functions use a balanced panel, 1995–96 through 2000–01.

SOURCE: Authors' calculations based upon data provided by the New York City Board of Education, the Ohio Department of Education, and Ohio Education Association.

Table 4. Comparison of percentile rankings by quantitative technique, Ohio schools: 1995–99

	EPF (mathematics)	DEA	APM (reading)	APM (mathematics)	Cost function
EPF (reading)	D2 = 0.1 D = 11.4 C = 76.8 U = 11.7 U2 = 0.0	D2 = 0.0 D = 13.4 C = 69.6 U = 16.3 U2 = 0.8	D2 = 0.0 D = 11.9 C = 76.3 U = 11.7 U2 = 0.1	D2 = 0.1 D = 14.4 C = 70.9 U = 14.4 U2 = 0.1	D2 = 0.8 D = 16.0 C = 66.6 U = 15.7 U2 = 0.9
EPF (mathematics)	†	D2 = 0.0 D = 12.5 C = 71.7 U = 14.8 U2 = 1.1	D2 = 0.1 D = 14.8 C = 70.4 U = 14.3 U2 = 0.4	D2 = 0.0 D = 10.8 C = 78.6 U = 10.5 U2 = 0.1	D2 = 0.8 D = 14.6 C = 69.1 U = 15.1 U2 = 0.5
DEA	†	†	D2 = 0.0 D = 16.0 C = 72.5 U = 11.5 U2 = 0.0	D2 = 0.4 D = 14.4 C = 74.4 U = 10.8 U2 = 0.0	D2 = 0.7 D = 15.6 C = 64.1 U = 15.1 U2 = 0.9
APM (reading)	†	†	†	D2 = 0.3 D = 12.7 C = 74.1 U = 12.7 U2 = 0.3	D2 = 1.1 D = 15.1 C = 68.0 U = 14.6 U2 = 1.3
APM (mathematics)	†	†	†	†	D2 = 0.7 D = 15.6 C = 67.8 U = 15.1 U2 = 0.9

† Not applicable.

NOTE: The schools are divided into the top group (schools whose efficiency is above the 90th percentile), the bottom group (schools whose efficiency is below the 10th percentile) and the middle group (schools whose efficiency ranges from the 10th to the 90th percentiles). In each cell, a series of numbers indicates the percentage of the schools that move, or do not move, from one of the percentile groups to another when the method listed in the column is used rather than the method listed in the row. D2 is the percentage of schools that move from top to bottom. D combines the schools that move from top to middle and the schools that move from middle to bottom. C is the percentage of schools that are ranked in the same percentile group according to both methods. U combines the schools that move from bottom to middle and the schools that move from middle to top. U2 designates the percentage of schools that move from bottom to top. Data for DEA and APMs are from 1997–98; EPFs and cost functions use a balanced panel, 1995–96 through 1998–99.

SOURCE: Authors' calculations based upon data provided by the New York City Board of Education, the Ohio Department of Education, and Ohio Education Association.

are in these two categories. None of the schools in this cell are in the U2 group ("up two"), that is, none of the schools move from the bottom to the top.

Table 4 presents results for Ohio and table 5 presents results for New York City. The least consistent ranking comparisons are those comparing the New York City DEA results to the other methods. As shown in table 5, a high proportion of schools are awarded a higher rank based upon the DEA measures than they are awarded based upon other measures. For example, the second cell in row 1 of table 5 indicates that 48.8 percent of schools are ranked one group higher using DEA as compared to the reading EPF and 5.3 percent of schools are ranked two groups higher. This pattern is the result of a large proportion of New York City schools being rated as fully efficient using DEA. Thus, since many schools earn 100 percent efficiency scores (tying for first place), the highest percentile grouping in the New York City DEA models actually includes more than 10 percent of schools.[9]

[9] Other specifications of the DEA model produce lower proportions of schools rated efficient (see Rubenstein 2003). For comparative purposes, the DEA specification presented in this paper uses the same combination of inputs and outputs as the other techniques.

Table 5. Comparison of percentile rankings by quantitative technique, New York City schools: 1995–2001

	EPF (mathematics)	DEA	APM (reading)	APM (mathematics)	Cost function
EPF (reading)	D2 = 0.0 D = 6.1 C = 87.7 U = 6.1 U2 = 0.0	D2 = 0.2 D = 10.8 C = 34.9 U = 48.8 U2 = 5.3	D2 = 0.0 D = 14.5 C = 71.4 U = 13.8 U2 = 0.3	D2 = 0.0 D = 14.8 C = 70.4 U = 14.8 U2 = 0.0	D2 = 1.0 D = 15.5 C = 66.6 U = 16.4 U2 = 0.5
EPF (mathematics)	†	D2 = 0.5 D = 9.1 C = 37.0 U = 48.2 U2 = 5.2	D2 = 0.2 D = 14.6 C = 70.6 U = 14.3 U2 = 0.3	D2 = 0.0 D = 13.5 C = 73.1 U = 13.5 U2 = 0.0	D2 = 1.2 D = 14.6 C = 67.8 U = 16.0 U2 = 0.5
DEA	†	†	D2 = 4.8 D = 50.5 C = 33.7 U = 10.1 U2 = 0.8	D2 = 5.8 D = 48.8 C = 33.7 U = 11.1 U2 = 0.5	D2 = 6.8 D = 46.9 C = 34.9 U = 10.8 U2 = 0.7
APM (reading)	†	†	†	D2 = 0.0 D = 11.6 C = 76.9 U = 11.3 U2 = 0.2	D2 = 1.5 D = 14.8 C = 66.9 U = 15.8 U2 = 1.0
APM (mathematics)	†	†	†	†	D2 = 0.5 D = 16.5 C = 66.8 U = 15.1 U2 = 1.2

† Not applicable.

NOTE: The schools are divided into the top group (schools whose efficiency is above the 90th percentile), the bottom group (schools whose efficiency is below the 10th percentile) and the middle group (schools whose efficiency ranges from the 10th to the 90th percentiles). In each cell, a series of numbers indicates the percentage of the schools that move, or do not move, from one of the percentile groups to another when the method listed in the column is used rather than the method listed in the row. D2 is the percentage of schools that move from top to bottom. D combines the schools that move from top to middle and the schools that move from middle to bottom. C is the percentage of schools that are ranked in the same percentile group according to both methods. U combines the schools that move from bottom to middle and the schools that move from middle to top. U2 designates the percentage of schools that move from bottom to top. Data for DEA and APMs are from 1999–2000; EPFs and cost functions use a balanced panel, 1995–96 through 2000–01.

SOURCE: Authors' calculations based upon data provided by the New York City Board of Education, the Ohio Department of Education, and Ohio Education Association.

If these methods are used to distribute rewards and sanctions to schools, it may be particularly distressing to find that schools are labeled as among the highest performers using one method or output measure and among the lowest performers with another. Overall, as a broad sweep, the results in tables 4 and 5 are somewhat surprising as they show relatively little movement between the top and bottom groups across methods, even ones as uncorrelated as cost functions and EPFs or ones as unrelated conceptually and empirically as EPFs and DEA. In general, the DEA measures show the most movement—again an interesting result because, while they are not empirically the least correlated with other methods, they are, arguably, the least related conceptually. That is, an EPF and a cost function are theoretically the inverse of one another, and an APM is an "atheoretical" EPF. So the EPFs, cost measures, and APMs are highly related in theory. But DEA, while an "input-output" type model, differs in its ability to choose frontier schools that excel in only one output.

As noted above, these comparisons raise the question, what are the characteristics of the schools that shift groups across methods? While providing a satisfying answer to this question is outside the scope of this paper, we compare the characteristics of the set of schools remaining in the same ranking category across two methods with those that changed categories. The results were intriguing, if only suggestive of questions for further investigation. As an example, in the New York City analyses, a series of pair-wise comparisons between the rankings based upon the different methods found that in 13 out of 15 of these comparisons, the schools with constant rankings were larger than the schools that had shifted categories, either up or down, and exhibited lower expenditures per pupil; in 14 out of 15 comparisons, the consistently ranked schools had a higher share of licensed teachers, experienced teachers, and teachers with master's degrees; and in all cases, the consistent schools had more teachers with at least 2 years in the school than did the schools that changed categories. Similarly, in Ohio, we found that in 13 out of 15 pair-wise comparisons, the consistently ranked schools were larger. Other differences were less dramatic, though. These simple comparisons appear to support findings from other work (for example, Kane and Staiger 2002) indicating that performance measures for small schools may be particularly susceptible to measurement error and random events. The other comparisons also suggest the need for further work to investigate the circumstances under which schools are persistently rated as high or low performing.

Conclusions

The four methods of school efficiency measurement we examine use different methodological approaches, but all are related conceptually by their connection to economic output/input theory. That is, each method implicitly treats schools as "firms" that convert a variety of inputs (resources, employees, students, etc.) into an array of outputs (typically some measure of student performance on tests, though measures such as graduation rates, attendance, and social outcomes could be added or substituted). The characterization of schools as "firms" does not imply that schools are, or should be, factory-like organizations or profit-making entities. It does, however, imply that we must try to identify the most effective strategies for accomplishing the most we can with increasingly scarce resources.

Using similar school-level data sets for two jurisdictions provides a unique opportunity to examine the characteristics of several different methods of school performance measurement, and to compare the stability of results using these multiple methods. The comparisons of efficiency rankings from these techniques indicate that the efficiency scores and efficiency rankings are moderately consistent, generally producing midrange Pearson and Spearman rank correlation coefficients. This result suggests that caution may be warranted in using these methods to distinguish *subtle* differences in school performance, however. While others have described large inconsistencies in school rankings across grades and subject matter exams (Kane and Staiger 2002) and across specifications (Clotfelter and Ladd 1996), our results show that different analytic methods may also produce different results, *even when using largely the same data and specifications*. While it may not be altogether surprising that methods using panel data produce different results from the methods using cross-sectional data (due to the different data used), our results indicate that the two methods using panel data (EPFs and cost functions) tend to produce very different results from each other in both samples. In the New York City sample, the methods using cross-sectional data also produced low correlations, while in the Ohio data they were relatively high (over 0.60).

Different analytic methods may also produce different results, even when using largely the same data and specifications.

The results also suggest that the various methods are unlikely to produce vastly different lists of the highest and lowest performing schools. If the purpose of the analysis is to identify consistently high-performing and low-performing schools, perhaps to study best practices or choose candidates for intervention, then the use of these multiple methods may provide a more reliable approach than the use of a single method. Our analyses suggest that these outlier listings will not change substantially across techniques. If the purpose of the analysis is to provide detailed relative rankings of schools, however, our analyses suggest that the results may be too sensitive to the quantitative techniques employed to produce reliable rankings throughout the distribution of schools.

Our analyses also highlight some of the potential benefits and drawbacks of each of the four methods:

- *Adjusted performance measures (APMs).* APMs are the most tractable of the four methods, imposing the least onerous data and analysis requirements. Their reliance on single output measures, though, requires consensus on the most appropriate measure—or composite of measures—to use for the analysis. Because of their relative ease of estimation, they may be the best suited of the four methods for construction of annual performance measures or reports.
- *Education production functions (EPFs).* While EPFs have considerably larger data requirements than APMs, they may be more effective for identifying persistent, rather than random or transitory, differences across schools. In school systems with relatively stable groups of schools, the EPF procedure may be a feasible approach for identifying consistent performance differences. At the same time, they may be of limited use in dynamic, rapidly changing school systems. Like APMs, they raise issues regarding the appropriate selection of output measures and grade levels.
- *Data envelopment analysis (DEA).* DEA has the distinct advantage of allowing multiple outputs and inputs simultaneously, permitting schools to focus on particular strengths. This type of specialization may not, however, be generally accepted by educators or families if, for example, schools achieve high test scores through high dropout rates. While the measures can be constructed with a single year of data, they require extensive data management and specialized software.
- *Cost functions*. Cost functions may be particularly useful for evaluating performance and efficiency in school systems facing severe financial constraints. Like EPF measures, they are likely to be relatively stable and effective for identifying persistent differences across schools. Like DEA, they include multiple outputs simultaneously. Despite these benefits, though, they are likely to impose the most prodigious data requirements and may be the least intuitive of the four methods.

The wide variation in the quality and quantity of inputs that each school faces, along with the variety of choices regarding outputs, makes it extremely difficult to validly and reliably identify schools that are making the most effective use of their resources and most efficiently achieving their goals. Ultimately, none of the measures we explore in this paper may be well suited for drawing the sharp distinctions between schools necessary for high-stakes accountability systems. Unfortunately, though, simplistic measures of school performance, which do not account for the complex environment of schooling, risk identifying the wrong schools as being exemplars of high performance or failures in need of interventions (see Rubenstein, Schwartz, and Stiefel 2003). This problem is particularly critical when the performance measures are used to distribute rewards and sanctions. While the rankings produced with the techniques in this paper may be somewhat volatile and are often complex, they may produce more valid measures of a school's contribution to student learning than do measures that do not attempt to mitigate the effects of student socioeconomic status on outcomes. Thus, we may face an unavoidable tradeoff between simplicity and validity in constructing such measures. The efficiency measures examined in this study are not simple, but may move us closer to accurately and reliably identifying those schools that are making the most effective use of their resources to educate their students.

We may face an unavoidable tradeoff between simplicity and validity in constructing efficiency measures.

References

Charnes, A., Cooper, W.W., Lewin, A.Y., and Seiford, L.M. (1994). The DEA Process, Usages and Interpretations. In A. Charnes, W.W. Cooper, A.Y. Lewin, and L.M. Seiford, (Eds.), *Data Envelopment Analysis: Theory, Methodology and Application* (pp. 425–435). Boston: Kluwer Academic.

Clotfelter, C.T., and Ladd, H.F. (1996). Recognizing and Rewarding Success in Public Schools. In H.F. Ladd, (Ed.), *Holding Schools Accountable, Performance-Based Reform in Education* (pp. 23–63). Washington, DC: The Brookings Institution.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F.D., and York, R. (1966). *Equality of Educational Opportunity.* Washington, DC: U.S. Government Printing Office.

Cullen, J.B., and Reback, R. (2002). *Tinkering Toward Accolades: School Gaming Under a Performance Accountability System.* Unpublished manuscript, University of Michigan, Ann Arbor, MI.

Figlio, D.N. (2003). *Testing, Crime and Punishment.* Unpublished manuscript, University of Florida, Gainesville, FL.

Figlio, D.N., and Winicki, J. (2002). *Food for Thought: The Effects of School Accountability Plans on School Nutrition.* (NBER Working Paper No. w9319). Cambridge, MA: NBER.

Hanushek, E.A., and Raymond, M.E. (2002). The Confusing World of Educational Accountability. *National Tax Journal, 54*(2): 365–384.

Kane, T.J., and Staiger, D.O. (2002). The Promise and Pitfalls of Using Imprecise School Accountability Measures. *Journal of Economic Perspectives*, *16*(4): 91–114.

Ladd, H.F. (2002) School-Based Educational Accountability Systems: The Promise and the Pitfalls. *National Tax Journal, 54*(2): 385–400.

Ladd, H.F., and Walsh, R.P. (2002) Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right. *Economics of Education Review, 21*(1): 1–17.

Linn, R.L., and Haug, C. (2002). Stability of School-Building Accountability Scores and Gains. *Educational Evaluation and Policy Analysis, 24*(1): 29–36.

Reid, T.R. (2002, December 26). Seven States Adopt Four-Day Week, The Washington Post, p. A03.

Rubenstein, R. (2003). *The Reliability of School Performance Measures Using Data Envelopment Analysis.* Unpublished manuscript.

Rubenstein, R.H., Schwartz, A.E, and Stiefel, L. (2003). Better Than Raw: A Guide to Measuring Organizational Performance With Adjusted Performance Measures, *Public Administration Review, 63*(5): 607–615.

Schwartz, A.E., Stiefel, L., and Bel Hadj Amor, H. (2003) Measuring School Performance Using Cost Functions, Paper prepared for the Annual Conference of the American Education Finance Association, Orlando, FL, March 2003.

Schwartz, A.E., and Zabel, J. (2003) *Measuring School Performance Using Education Production Functions.* Paper prepared for the Annual Conference of the American Education Finance Association, Orlando, FL, March 2003.

Stiefel, L., Schwartz, A.E., Bel Hadj Amor, H., and Kim, D.Y. (2003). *Measuring School Performance Using Adjusted Performance Methods,* Paper prepared for the Annual Conference of the American Education Finance Association, Orlando, FL, March 2003.

Stiefel, L., Rubenstein, R.H., and Schwartz, A.E. (1999). Measuring School Efficiency Using School-Level Data: Theory and Practice. In A. Odden and M.E. Goertz, (Eds.), *School-Based Financing* (pp. 67–87). Thousand Oaks, CA: Corwin Press.

Court-Mandated Change: An Evaluation of the Efficacy of State Adequacy and Equity Indicators

Jennifer Park
Ronald A. Skinner
Education Week

About the Authors

Jennifer Park is a research associate with *Education Week.* She earned a B.A. in political science from Syracuse University, and an M.A. in education policy from Ohio State University. Her current areas of interest include state-level education policy and school finance. She can be contacted at jpark@epe.org.

Ron Skinner is the research director at *Education Week.* Ron has undergraduate degrees in political science and communications and a master's degree in public policy from the University of Central Florida. His research interests include state-level education policy and issues related to school climate. He can be contacted at rskinner@epe.org.

Court-Mandated Change: An Evaluation of the Efficacy of State Adequacy and Equity Indicators

Jennifer Park
Ronald A. Skinner
Education Week

Every January, *Education Week* releases its annual report, *Quality Counts*, which grades the states in several areas including the equity and adequacy of resources dedicated to education. The grades are based on a series of indicators developed and revised over the last few years with the advice of experts in the area of school finance. Data, mostly from the National Center for Education Statistics and the U.S. Census Bureau, are used to calculate the indices with the school district as the unit of analysis. As part of the effort to continually ensure the value of the data presented in *Quality Counts*, this paper will take advantage of recent court decisions mandating drastic changes in state education finance systems to evaluate the efficacy of the report's finance adequacy and equity measures.

In the four states studied for this paper—New Hampshire, New Jersey, Vermont, and Wyoming—state legislatures have implemented changes in the way the state collects and distributes money for education. These states were chosen for this analysis because, first, the impetus of a court decision in favor of the plaintiff forced the legislatures in these states to make sweeping changes to their education finance systems. Since these court-mandated changes tend to be more comprehensive, they should be more easily detected in data analyzed from the years before and after reforms were implemented. Second, these states were chosen for the relatively settled nature of these cases, meaning that a court decision and the necessary legislative action have taken place—regardless of whether all groups are happy with the outcomes. Finally, the timing of the court decision in these four states means that the federal data used to calculate these measures are available.

This paper begins by describing the finance measures used in *Quality Counts,* as well as the federal data used to calculate these measures. The paper then explains the litigation history of the court cases that mandated the changes to the education funding systems in each of the four states, and when these reforms were implemented. Next, for each state, based on the court case and school finance reform information, outcomes that can be expected from the data analysis are listed. Finally, this study analyzes the equity and adequacy measures calculated from data before and after states implemented finance reforms, to see if these indicators and the data they represent are accurately measuring school finance changes in the states.

Data Sources

The data used for the analysis described in this study were compiled from a number of sources and then merged into a single database to create variables for each school district in the nation. The following data sources and variables were used:

- U.S. Census Bureau, Public Elementary/Secondary Education Finance Data (commonly known as the National Center for Education Statistics' Common Core of Data Local Education Agency [School District] Finance Survey [F-33] Data), 1994–2000
  - State Identification Number (STATE)
  - School System Name (NAME)
  - School Level Code (SCHLEV)
  - NCES ID Code (NCESID)
  - Year of Data (YRDAT)
  - Fall Membership, October (V33)
  - Total Revenue From State Sources (TSTREV)[1]
  - Total Revenue From Local Sources (LOCREV)
  - Total Current Spending for Elementary/Secondary Programs (TCURELSC)
- National Center for Education Statistics, Common Core of Data, Public Elementary/Secondary School District Universe Data, 1994–2000
  - NCES Agency ID (LEAID)
  - State Abbreviation (ST##)
  - Agency Type Code (TYPE##)
  - Total Schools (SCH##)
  - Students with an Individualized Education Plan (SPECED##)
- Chambers Cost of Education Index (1993–94)
  - NCES Agency ID (NLEA_ID)
  - Cost of Education Index (CEIL93)
- U.S. Census Bureau, Small Area Income and Poverty Estimates, School District Estimates, 1995 and 1997
  - FIPS State Code (FIPS)
  - CCD District ID (CCDID)
  - District Name (DISTNAME)
  - Estimated Total Population (TOTALPOP)
  - Estimated Population of Children 5 to 17 Years of Age (CHILD)
  - Estimated Number of Poor Children 5 to 17 Years of Age Who Are Related to the Head of the Household (POORCHRN)
- School District Data Book, 1990 Census School District Special Tabulation, U.S. Summary
  - Table H062—Aggregate Value of Specified Owner-Occupied Housing Units by Mortgage Status (WEALTH)
  - Table P202—Area in Square Kilometers (AREA)
- National Center for Education Statistics, Common Core of Data, *Early Estimates of Public Elementary and Secondary School Education Statistics,* school year 1996–97 to school year 2001–02.
  - Table 7: Per pupil expenditure

These data were compiled for every district in the nation, and merged into one file using the NCES district code in each data set as the unique identifier. Several variables were calculated from these data to create the equity and adequacy indicators discussed in this study (see the appendix for a detailed description of these variables and calculations).

Equity and Adequacy Indicators

Each year in its report *Quality Counts, Education Week* uses the most recent data available from the sources listed above to grade states on how adequately and equitably they fund education. This study uses some of the same indicators used in the grading process for *Quality Counts*. For equity, these measures are state equalization effort, targeting score, wealth-neutrality score, coefficient of variation, and McLoone Index. For adequacy, the indicators used in this study are adequacy index, education spending per student (adjusted for regional cost differences), and the percent of students in districts with per pupil expenditures at or above the U.S. average.

[1] This variable represents all state revenue received by each district including revenue from general formula aid, categorical programs, and all other revenues from the state.

In the grading process for *Quality Counts*, state equalization effort accounts for 50 percent of the equity grade, wealth-neutrality for 25 percent, and coefficient of variation and McLoone Index each count for 12.5 percent of the grade. For adequacy, education spending per student and the adequacy index each count for 40 percent of the grade.[2] Following is a description of each of the measures used in this analysis.

State equalization effort

This indicator is based on the concept that states can help equalize funding across districts in two ways: by providing all or most of the total funding so there are no discrepancies across districts, or by targeting more revenue to property-poor districts that are not able to raise as much revenue locally. Most states use a combination of these two strategies. The state equalization effort indicator measures these two approaches and is the state share of total state and local funding adjusted by the degree to which these funds are targeted to poorer districts.

The score for state equalization effort depends on both a targeting score and the percentage of state funding. The *targeting score* represents the extent to which state funds are targeted to property-poor districts. In *Quality Counts 2002*, targeting score values ranged from –.53 to zero, where the more negative a value, the more state funds are targeted to poorer districts. A targeting score of zero means that the state is not targeting funds to property-poor districts.

The targeting score was calculated using multiple regression. The regression model was designed to control for other district characteristics, besides wealth, that could influence state aid. The dependent variable in the model was adjusted state revenue per pupil. This variable was adjusted to reflect geographic cost differences relative to each state, and was also indexed so that the state's average per pupil figure was 1. The independent variables in the model include adjusted district wealth per pupil,[3] percent of students in poverty, percent of children in special education (i.e., those with individualized education plans), student enrollment, and land area per pupil (all indexed to the state average). The coefficient for the first independent variable (the index of adjusted state revenue per pupil) from the regression serves as the targeting score.

State equalization effort is the state's share of funding multiplied by the inverse of this targeting score. For example, in 2000, state aid in Florida accounted for 54.7 percent of total (state and local) revenue, which was below the national average for that year (57 percent). Florida, however, had a targeting score of –.473, meaning it targeted more funds to property-poor districts. Therefore, its effort to equalize funding was higher than what the state share of funding would suggest. The calculation for Florida's state equalization effort for 2000 is as follows:

State equalization effort = State share of funding x (1 – targeting score)
= 54.7 percent x (1 – (–.473))
= 80.6 percent

The state equalization effort adjusts the state share of funding to reflect the effort the state has made to target funds to property-poor districts. If the state's targeting score is zero, the state equalization effort will be the same as its state share of funding. In *Quality Counts 2003*, state equalization effort values ranged from 43 percent to 98 percent.

Wealth-neutrality score

Like the targeting score, the wealth-neutrality score also shows the degree to which revenue is related to the property wealth of districts. However, this indicator considers both state and local revenue. West Virginia, for example, had a targeting score of –.039 in 2000, indicating that the state is targeting aid to property-poor districts. When local revenue was also considered in the wealth-neutrality score, however, West Virginia had a .084, meaning that higher property wealth is still linked to more revenue.

2 The remaining 20 percent consisted of measures not used for this analysis—taxable resources spent on education (15 percent) and average annual rate of change in expenditures per pupil (5 percent).

3 The use of wealth versus income in this model was tested with a sensitivity analysis when this indicator was first introduced to Quality Counts in 2000. In a comparison of R^2s for each state using either wealth or income in the model, wealth often explained more variance in state aid than income.

The wealth-neutrality score was also calculated using regression. The dependent variable in the model was adjusted state and local revenue per weighted pupil. The variable was adjusted to reflect geographic cost differences relative to each state, and weighted by student needs (i.e., poor students = 1.2, and special education students = 2.3). The figure was also indexed so that each state's average per pupil figure was 1. The single independent variable in the model was adjusted property wealth per weighted pupil, also adjusted to reflect cost differences and student needs, and indexed to the state average. The coefficient for the independent variable (adjusted property wealth per weighted pupil) from the regression serves as the wealth-neutrality score.

In *Quality Counts 2003*, wealth-neutrality scores ranged from –.189 to .311. A negative score means that, on average, property-poor districts actually have more funding per weighted pupil than wealthy districts do, and a positive score means the opposite, that wealthy districts have more funding per weighted pupil than property-poor districts do. Only eleven states had a negative wealth-neutrality score in *Quality Counts 2003*.

McLoone Index

The McLoone Index is based on the assumption that if all students in the state were lined up according to the amount their districts spent on them, perfect equity would be achieved if every district spent at least as much as was spent on the pupil in the middle of the distribution, or the median. The McLoone Index is the ratio of the total amount spent on pupils below the median to the amount that would need to be spent to raise all students to the median.

The McLoone Index was calculated by first computing the median-level expenditure per pupil for each state (adjusted to reflect cost differences and student needs). The second calculation was the total number of dollars spent on students whose per pupil expenditure was below the median. Finally, that figure was divided by the total amount that would be spent if every pupil below the median had the median-level expenditure.

For example, the median-level expenditure per pupil (adjusted to reflect student needs) in Indiana for *Quality Counts 2003* was approximately $5,583. The total amount spent on students who were below that mark was about $3.04 billion. In order to spend $5,583 on each of those pupils below the median, the state would need to spend $3.32 billion. The calculation for Indiana's McLoone Index for 2000 is as follows:

McLoone Index
= Amount spent on pupils below the median / Amount needed to be spent to achieve "equity"
= $3.04 billion / $3.32 billion
= 91.56 percent

This indicates that state and local spending on children below the median was about 92 percent of what was needed in 2000 to raise all students to the median expenditure. McLoone Index values in *Quality Counts 2003* ranged from 87 percent to 100 percent, where perfect equity is represented by 100 percent and the greatest inequity by 0 percent.

Coefficient of variation

The coefficient of variation is a measure of the discrepancy in funding across the districts in a state. This measure was calculated by dividing the standard deviation of adjusted spending per weighted pupil (adjusted to reflect cost differences and student needs) by the state's average spending per pupil. For example, the standard deviation for spending in Maryland in 2000 was about $584. The average spending per pupil for Maryland for the same year was $6,265. The calculation for Maryland's coefficient of variation in 2000 is as follows:

Coefficient of variation
= Standard deviation of adjusted spending per weighted pupil / Average spending per pupil
= $584 / $6,265
= 9.3 percent

If all districts in a state spent exactly the same amount per pupil, its coefficient of variation would be zero. As the coefficient gets higher, it means the variation in the amounts spent across districts also gets higher. As the coefficient gets lower, it indicates greater equity. In *Quality Counts 2003*, the range of values for the coefficient of variation was 6 percent to 32 percent.

Adequacy index

Since there is no consensus about how much money is necessary to provide an "adequate" education, the adequacy index uses the national average as the benchmark against which to gauge state spending. While it may seem intuitive to measure adequacy simply by calculating the percent of students in districts where spending eclipses the national average, that calculation is not ideal. Imagine if every district in a state were to spend exactly $5,593 per student, just $1 below the 2000 national average. Spending on every student would be amazingly close to what is considered to be adequate, yet no student in the state would seem to be enrolled in a district with adequate funding. The adequacy index takes into account both the number (or percentage) of students enrolled in districts with adequate spending, and the degree to which spending is below adequate in districts where per pupil expenditures are below the national average.

The adequacy index was calculated using district-level spending that was adjusted for student needs and regional cost differences. Each district where the per pupil spending was equal to or exceeded the national average received a score of 1 times the number of students in the district. Districts where the adjusted spending per pupil was below the national average received a score equal to their per pupil spending divided by the national average and then multiplied by the number of pupils in the district. The adequacy index is the sum of district scores divided by the total number of students in the state. If all districts spent above the U.S. average, the state received a perfect index of 100.

Example:

District	Enrollment	Per pupil spending
1	400	$7,000
2	450	$6,000
3	500	$5,000
4	300	$4,000
5	350	$3,000
Total	2,000	

In the above example, districts 1 and 2 are the only ones providing an adequate education (i.e., equal to or above the 2000 national average, $5,594). Scores for these districts are equal to their student enrollment. The percent of students attending schools in districts with adequate spending, then, is 850 divided by 2,000, or 42.5 percent. This is the equivalent of the indicator *percent of students in districts with per pupil expenditures at or above the U.S. average.* This figure, however, does not account for how close spending is to adequate in the remaining three districts, a problem that is corrected in the calculations below.

District	Score
1	400
2	450

Districts 3 through 5 are below the U.S. average, so assigning scores to each district will tell us how "far" they are from adequate spending. Their scores are equal to their average spending divided by the U.S. average and multiplied by the number of pupils in the district, as shown below.

District	Score
3	446.91 = ($5,000 / $5,594) * 500
4	214.52 = ($4,000 / $5,594) * 300
5	187.70 = ($3,000 / $5,594) * 350
Total	1,699.13 (for all five districts)
Adequacy index	= 1,699.13 / 2,000
	= 84.96

This value represents an index against which it is possible to compare the relative adequacy of the 50 states and the District of Columbia. In *Quality Counts 2003*, values for the adequacy index ranged from 70 to 100.

Education spending per student

For this indicator, each state's education spending per student was based on per pupil expenditure data taken from the NCES report, *Early Estimates of Public Elementary and Secondary School Education Statistics*. With the Chambers Cost-of-Education Index, each state's per pupil expenditure was adjusted for regional cost differences by dividing the expenditure by the state's figure from the cost-of-education index.

Methodology

The purpose of this study was to test the extent to which the equity and adequacy indicators used in *Quality Counts* each year represent actual changes in the way states collect and distribute funds for education. Four states were chosen as the sample for this analysis based on a recent court decision in each state and a subsequent change in the education finance system. Information regarding the school finance and litigation history of these four states—New Hampshire, New Jersey, Vermont, and Wyoming—was collected from a variety of sources. These included court case decisions, legislation on state funding system changes, analyses from private groups, the NCES report *Public School Finance Programs of the U.S. and Canada: 1998–99* (National Center for Education Statistics 2001), and other sources.

The equity and adequacy indicators used in *Quality Counts* were calculated for each state for the years before and after education finance reform occurred in the states. The data for these indicators came in part from previously published data in past issues of *Quality Counts*, and for those years that were not covered in past issues, an additional analysis was conducted for this paper.

The equity and adequacy indicators used in* Quality Counts *were calculated for each state for the years before and after education finance reform occurred in the states.

Data were available for all equity indicators other than the coefficient of variation for the years 1996, 1997, 1999, and 2000 from past *Quality Counts* publications. Data for 1994, 1995, and 1998 were calculated for this paper to better detect trends over time. Since coefficient of variation has always been used in *Quality Counts* as a measure of equity, data for this indicator were available from past issues for the years 1994–1997, 1999, and 2000.

For adequacy, there were fewer years of data available from past publications since *Education Week* only recently started calculating the adequacy index. Results for the adequacy index and the percent of students in districts with per pupil expenditures at or above the U.S. average were only available for 1999 and 2000 from previous issues. Additional analyses of these indices were conducted for this paper for 1994, 1995, and 1996 data. Like the coefficient of variation, education spending per student is a measure that has always been used in *Quality Counts*' adequacy grading, so data were available from 1996 to 2002. This indicator is calculated from more recent data, as it is based on the NCES "Early Estimates" report. The most recent data available for the equity indicators in this study, and the data used for the adequacy index and percent of students in districts with per pupil expenditures at or above the U.S. average is from the 1999–2000 school year.

State Finance and Litigation History

New Hampshire

Prior to the 1999–2000 school year, the funding system for public education in New Hampshire relied heavily on local property taxes. On average, 90 percent of funding came from local property taxes, 7 percent from the state, and 3 percent from the federal government. Under the old system, school districts set the annual budgets for their schools. Once the voters approved the budget, the budget was sent to the state, and the state determined the appropriate property tax necessary to raise the funds. In New Hampshire, local school boards do not have the power to levy taxes. While the state has tried at least twice before to revise its funding formula to erase disparities across districts (in 1919 and 1947) by setting maximum property tax rates above which the state would provide the necessary support, on both occasions the legislature has failed to provide the necessary funds (National Center for Education Statistics 2001).

New Hampshire's system of state funding that was contested in court was based on a foundation formula that included weights for special education, vocational education, and grade-level enrollments. Local fiscal capacity was measured by assessed property valuation, school tax rates, and personal income. Although the state intended to fund the average district (based on wealth) at 8 percent of operating expenditures, every year the appropriated funds were less than what was needed. The state had a few categorical programs in this old system including special education, vocational education, the Kindergarten Incentive program, teacher retirement and benefits, and capital outlay and debt service (National Center for Education Statistics 2001).

In a series of rulings (Claremont I, II, and III), the New Hampshire Supreme Court mandated changes to this education funding system. Most notably, the *Claremont II* ruling declared the state's system unconstitutional and ordered that the system may not remain in effect beyond the 1998–99 school year. After much legislative wrangling, the system that passed the legislature in April 1999 under this mandate created a statewide property tax for education, and raised business taxes to cover the additional costs of providing an adequate education (Viadero 2001). A new system of distributing funds for education was implemented in the 1999–2000 school year. The process for distributing funds in this new system was relatively similar to the old system; the major change was the collection of revenues through the statewide property tax. The state also greatly increased its responsibility and share of education funding by implementing a base cost in the foundation formula, which was set by analyzing the expenditures of a select group of schools. This base cost was the average per pupil expenditure of the lowest spending half of elementary schools in districts where 40 to 60 percent of students scored at or above *Basic*[4] on the New Hampshire Educational Improvement and Assessment Program (National Center for Education Statistics 2001).

An analysis of tax rates and school spending in New Hampshire concluded that disparities in spending per student across districts did not change greatly.

An analysis of tax rates and school spending conducted by the New Hampshire Center for Public Policy Studies looked at the differences in actual spending in the state before and after the implementation of the new law. According to the study *School Finance Reform: The First Two Years,* while the new legislation did have an impact on the amount spent overall on education in New Hampshire, the study concluded that disparities in spending per student across districts did not change greatly (Hall 2002).

Based on the reforms implemented, the equity and adequacy data for New Hampshire should experience the following changes:

- The overall assessment of equity in the state should remain fairly constant before and after the 1998–99 school year. Taxation under the new system has been more equitable, but the distribution of spending for education has remained fairly constant. The main indicator that should change is the state equalization effort since the state greatly increased its share of funding.
- New Hampshire's wealth-neutrality score should also rise slightly due to the switch to a statewide property tax.
- In terms of adequacy, the state should improve on the adequacy index and education spending per pupil after the 1998–99 school year, since more money was provided across the board for education.

Because the new education finance plan was approved in the late spring of 1999, many budgets for the 1999–2000 school year had already been passed. More changes in the adequacy indices should be reflected in the 2000–01 data, as school boards were able to pass budgets with full knowledge of the new system.

New Jersey

The system for funding public education in New Jersey was first declared unconstitutional in 1973 on grounds that it did not meet the "thorough and efficient" clause of the state constitution, in 1973 (Robinson v. Cahill, 303 A.2d 273). "Since that decision, the supreme court has issued over a dozen school finance opinions, the latest in May 2002," (Advocacy Center for Children's Educational Success With Standards 2002). By far the most well-known series of decisions come from the case of *Abbott v. Burke,* named for Raymond Abbott, an elementary student on whose behalf the first suit was filed in 1981, against

[4] Two scoring levels, *Proficient* and *Advanced*, were above *Basic*, and one scoring level, *Novice*, was below *Basic*.

then state education commissioner Fred Burke. Nine years later, the first state supreme court ruling in the case declared that the funding system for poor, urban districts was inadequate. One unique aspect to this case is that the court addressed only the poorest districts in the state (Newman 1990).

Before the court required the state to change its education funding system again in 1997, New Jersey state education aid came from the general fund (20.1 percent), and the property tax relief fund, which was revenue from a state income tax (79.9 percent). Property taxes were the only source of local funds for schools, exacerbating inequalities in wealth at the local level. State aid was distributed with a foundation formula that included weights for grade level, vocational school, and adult education enrollments, with district shares also weighted for property wealth and aggregate income. Additionally, in response to the 1990 New Jersey Supreme Court ruling, the state was already making adjustments for 30 special-needs districts, where these districts had a foundation level that was 5 percent higher than other districts in the state, a different calculation for local fiscal capacity, and different budget cap rules. New Jersey also had several categorical programs including transportation, capital outlay and debt service, teacher retirement, special education, compensatory education, and private school aid (Gold, Smith, and Lawton 1995).

The most prominent feature of the finance reform efforts in New Jersey is the court mandated requirement for the state to fund the 30 poorest districts in the state.

In the years following the first *Abbott* ruling at least three different finance plans were implemented in New Jersey, ending with the Comprehensive Educational Improvement and Financing Act of 1996 (National Center for Education Statistics 2001). This act was first implemented in the 1997–98 school year and remains in place, although the details of provisions within the act have been continually litigated. By far, the most prominent feature of the finance reform efforts in New Jersey is the court mandated requirement for the state to fund the 30 poorest districts in the state, known as the *Abbott* districts, at the same level as the state's wealthiest districts. This "Parity Remedy Aid" was ordered in the fourth *Abbott v. Burke* decision, in 1997.

New Jersey's reformed system for ensuring equality between these poorest districts and the wealthiest in the state was made up mostly of supplemental and parity aid to these districts. The state still used a foundation formula to distribute its core curriculum standards aid. Pupil counts were used as the basis for this formula and were weighted by instructional levels. A "T&E" (thorough and efficient) budget amount was the basis of this formula, and was adjusted for inflation every 2 years. Additional aid was provided for the *Abbott* districts. New Jersey also added a categorical program for early childhood education in its new system, which was targeted to low-income districts (National Center for Education Statistics 2001).

Based on the changes in the state resulting from *Abbott v. Burke,* the data from these indicators should change in the following ways:

- Equity indicators should have a clear increase for the state after the 1997–98 school year, since many of the reforms were targeted at bringing the *Abbott* districts on par with wealthier districts. This should mostly be evident in the state's targeting score.
- For adequacy, the indicators should improve slightly since the state increased funding after the 1997–98 school year.

Vermont

Vermont has had a fluctuating investment in its education system over the last 40 years. According to a report from the National Center for Education Statistics, "between 1964 and 1997, the state share of basic educational expenses varied between 20% and 37%" (National Center for Education Statistics 2001). The report describes a pattern of state funding during that time in which Vermont would take legislative action to reform the finance formula and raise funding when the state share of funding dropped to around 20 percent. The state would gradually allow the state share to drop toward 20 percent again, and then take new action. By 1997, the year a court case prompted the most recent response from the state, the state share of funding was about 25.3 percent (National Center for Education Statistics 2001).

The Vermont system for financing education that was ultimately ruled unconstitutional was based on a foundation formula. The formula was allocated on average daily membership measured over 2 years. Local fiscal capacity was based on property value and income, although income was only a small factor in the formula. Average daily membership counts were weighted for secondary student enrollment (1.25), poverty (1.25), and transportation costs (1.0384 to 1.0714). The state foundation cost was $4,025, with no aid going to "gold towns" that were able to raise 1.5 times more than this level with local resources. There were no minimum or maximum expenditure limits; local voter willingness to pay was the only upper limit on these towns. The old system included several categorical programs including transportation, capital outlay and debt service, teacher retirement, special education, vocational education, private school aid, and early childhood education (Gold, Smith, and Lawton 1995).

The Vermont Supreme Court ruled that spending in a locality should not be a function of property wealth.

In *Brigham v. State,* 692 A.2d 384 (Vt. 1997), the Vermont Supreme Court ruled that the state foundation aid program and the state system of relying heavily on locally raised revenues were not in line with the requirements of the Vermont constitution. In its decision the court wrote that:

> In Vermont the right to education is so integral to our constitutional form of government, and its guarantees of political and civil rights, that any statutory framework that infringes upon the equal enjoyment of the right bears a commensurate heavy burden of justification. (Brigham v. State, 692 A.2d 384 at 5 [Vt. 1997])

While the decision indicated that equality in per pupil spending across the state was not a necessary remedy for the state's education finance system, it also declared that spending in a locality should not be a function of property wealth.

Only 4 months after the ruling, the Vermont legislature passed the Equal Education Opportunity Act of 1997, No. 60 of the Acts of 1997 (Act 60), which created an income-sensitive, statewide property tax. Under this system, most residents actually paid a state tax of 2 percent of their income, but wealthier residents and those living in residences on more than 2 acres of land also continued to pay a 1.1 percent tax on their property value (Heaps and Woolf 1997).[5] Revenues from this state property tax were distributed to the districts at a rate of $5,448 starting in 1997, with the new legislation requiring an annual increase in the per pupil funding matching the most recent cumulative price index (Tit. 16 § 4011 [2003]). Prior to Act 60, some districts spent less per pupil than what came directly from the state after Act 60, yet the median spending per pupil in the state before 1997 was closer to $6,200 (Heaps and Woolf 1997).

In order to spend more than the amount offered by the state, towns had to implement their own local property taxes, and under Act 60, these taxes also had to be income sensitive. In order to meet the court-mandated requirement that school spending not be related to town property wealth, any local funds generated by a local property tax were subject to a state-imposed equalized yield. This ensured that a property-poor town levying a 5 percent local property tax would get the same additional amount per pupil as a property-rich town taxing at the same rate, even if the property-rich town generated significantly more revenue.

An important point about the Act 60 reforms relative to this analysis is that the reforms were phased in over 3 years, so the act was not fully in place until 2001; however, the major reforms occurred in 1999 (National

[5] Vermont has since revised the school finance system that was implemented following Act 60. The changes, which will take effect in the 2004–05 school year, reduce the reliance on property taxes by raising the sales tax rate, eliminate the sharing pool, and increase per pupil aid to schools.

Center for Education Statistics 2001). In 1999, Vermont changed its foundation formula to a two-tier system and implemented a statewide property tax. The two-tier formula included a block grant as the first tier, and a guaranteed-yield program with a recapture provision as the second. The new formula was based on equalized pupils weighted by grade, poverty (measured by food stamp participation), and enrollment of English Language Learners. The formula also adjusted for small schools, small school enrollment stability, and whether a town was a receiving or sending town based on the recapture provision (National Center for Education Statistics 2001).

An early analysis of state education spending data by William Mathis at the University of Vermont found that disparities in spending per pupil diminished across the state, although not as much as the disparities in property tax rates (Mathis 2000). Vermont continued to make progress toward the goal of reducing disparities in spending, according to a February 2002 report from the Rural School and Community Trust (Jimerson 2002). This report found that the difference in spending between property-poor and property-rich towns was about 37 percent, or $2,100 per pupil, in fiscal year (FY) 1998 compared to a difference of only 13 percent, or $900 per pupil, in FY 2002.

Taxpayer reaction in the wealthiest towns in Vermont was for residents to forego raising local taxes, and instead develop charitable funds to give gifts to local schools.

Based on the reforms implemented from Act 60, the equity and adequacy data should experience the following changes:

- Adequacy index, average education spending per pupil, and the percent of students in districts spending at or above the U.S. average should improve slightly in the 1998–99 school year as students in the poorer districts receive more funding.
- In terms of equity, the state share of funding and targeting scores should increase after 1998, and the McLoone Index should also improve as the spending in the poorest districts rises to at least the level of the state grant.
- Wealth neutrality should improve slightly with the statewide property tax and recapture provision.

An important aspect of the new system implemented with Act 60 was taxpayer reaction in the wealthiest towns in Vermont. Residents in these towns experienced the largest tax increases, and lost a large share of the money raised by these taxes to less fortunate towns. One response by these towns was for residents to forego raising local taxes, and instead develop charitable funds to give gifts to local schools, thus avoiding the recapture provision of Act 60. Mathis points out that "16 towns raised $7.3 million in gifts and thereby denied $15.7 million in recaptured funds" (Mathis 2000). It is important to note that if these funds are not included in reports to the federal government regarding state funding they will not be reflected in the indicators used in *Quality Counts.*

Wyoming

Some of the first signs of trouble for school finance in Wyoming date back to 1980, when the Wyoming Supreme Court first ruled the system unconstitutional because it failed to offer equal protection as the state constitution mandates. By 1983, the state had implemented reforms that required minimum local taxes and created a recapture feature to take money from wealthier districts to help support smaller, rural districts. While these changes were supposed to be a temporary fix, the system remained in place (Miller 1995).

Before the Wyoming Supreme Court required changes to the finance system again in 1995, the state used two main funding sources: a 12-mill statewide property tax and mineral production royalties from the federal government. The state system was based on a formula allocated by classroom units where local capacity was assessed mainly through property valuation. The formula also had a recapture provision, so if a local district's revenue was greater than 109 percent of the state minimum level then the district had to return those funds to the state. Wyoming had only two main categorical programs, special education and transportation (Gold, Smith, and Lawton 1995).

In 1993, four large districts sued the state, and the 1995 *Campbell* decision required the state to:

> (1) define the 'basket' of education every Wyoming child should receive—the best we can do, not just a minimal education; (2) undertake the cost of education studies to determine the actual cost of providing the basket in the various sizes and types of school districts, taking into account the needs of different kinds of students; and (3) fund the basket—in that order. (National Center for Education Statistics 2001)

Thus, the court required not just a remedy to the equity problem in the state, but it also answered the question of what is adequate.

The state hired an outside contractor to conduct the necessary research. By April of 1997, Management Analysis & Planning Associates submitted their proposal to the Wyoming legislature. Among the changes suggested was a move from a formula based on classrooms, a model which favored smaller, more rural districts, to a model built around average daily membership. The new plan also added cost adjustments for cost-of-living, teacher seniority, and pupil characteristics, and it made the state the authority for determining total district revenue and education-related taxes (Guthrie et al. 1997).

Most new provisions for Wyoming's reformed system were put in place for the 1998–99 school year. The new school aid formula was mostly a block grant based on average daily membership measured over 3 years, with adjustments for local cost of living and districts with less than 1,350 average daily membership. Local capacity was still measured by local average property valuation, but the recapture provision was reduced from districts that raised 109 percent of the state minimum to those able to raise 100 percent. The state developed prototypes based on enrollment and class-size levels to define the educational basket of costs. This basket was built around 25 cost components in five categories: personnel, supplies, special services, special students characteristics, and district or regional characteristics. Wyoming still had only two main categorical programs for special education and transportation.

Based on the changes resulting from the *Campbell* case and the ensuing Cost-Based Block Grant model implemented by the legislature and phased in by the state during the 1997–98 and 1998–99 school years

- Adequacy in the state should improve slightly after the 1998–99 school year due to an additional $50 million necessary to enact the recommended reform on top of the $600 million the state already spends annually.
- Funding in Wyoming should become more equitable over this time, because the state now largely has control over total spending per student, and because the distribution of funding by the state has moved from a per classroom basis to a per student formula, the same unit of measure used in the equity indicators.

New Hampshire always had the lowest state equalization effort of the 50 states.

State Results

New Hampshire

Until this past year, New Hampshire had been one of the worst scoring states on the equity grades in *Quality Counts*. Over the last 4 years the state consistently received an F on equity grades. The main reason for this was the state's low share of education funding, and because of that, its low state equalization effort. New Hampshire always had the lowest state equalization effort of the 50 states, with a score in the low teens (table 1). On the other hand, New Hampshire had a very good targeting score; at one point, in 1997, it was –.734. According to these equity indicators, although New Hampshire may have had very little state funding for education, the state targeted what funding it did provide very heavily to property-poor districts.

In 2000, New Hampshire's state equalization effort skyrocketed to 57.1 percent. This is a very drastic change from 14.2 percent the year before and indicates that the state greatly increased its investment in education. Another interesting change in the equity indicators from 1999 to 2000 was a large jump in

Table 1. Changes in equity and adequacy indicators over time in New Hampshire

	1994	1995	1996	1997	1998	1999	2000	2001	2002
State equalization effort	12.8	11.2	10.8	13.0	14.1	14.2	57.1	—	—
Targeting score	–.608	–.584	–.537	–.734	–.547	–.558	–.009	—	—
Wealth-neutrality score	.161	.174	.152	.233	.174	.173	.162	—	—
McLoone Index	.864	.860	.878	.883	.876	.894	.887	—	—
Coefficient of variation	17.1	16.8	17.5	16.9	—	18.2	17.5	—	—
Education spending per student (adjusted for regional cost differences)	—	—	$5,541	$5,805	$5,942	$6,195	$6,437	$6,967	$7,563
Adequacy index	87.03	87.18	84.76	—	—	91.47	91.78	—	—
Percent of students in districts with per pupil expenditures at or above U.S. average	24.81	23.0	20.27	—	—	38.91	39.55	—	—

—Not available.

SOURCE: Bureau of the Census: Public Elementary/Secondary Education Finance Data, 1994–2000; Small Area Income and Poverty Estimates, School District Estimates, 1995 and 1997; and School District Data Book, 1990. National Center for Education Statistics: Common Core of Data: Public Elementary/Secondary School District Universe Data, 1994–2000, and *Early Estimates of Public Elementary and Secondary School Education Statistics,* school year 1996–97 to school year 2001–02; and Chambers Cost of Education Index, 1993–94.

New Hampshire's targeting score. In 2000, it was almost zero (–.009), indicating almost no targeting of funds. This is very significant, considering how heavily the state targeted funds to property-poor districts in the past. It appears that although the state made great strides to increase its funding overall, it no longer targeted funds as heavily to property-poor districts.

One change that would be expected with such an increase in state funding that did not coincide with the increases in the state's equalization effort and targeting score is an increase in New Hampshire's adequacy index. Since the state had such a strong increase in its share of funding, it would be expected that the state would improve on the adequacy index. Instead, the score for this index only rose slightly in 1999 and 2000 (table 2). When looking at education spending per student, however, it is clear that in 2001 and 2002, funding in New Hampshire grew substantially. From 1997 to 1999, education spending per student only increased by $390, while from 2000 to 2002, the amount of spending per student increased by over $1,000 per pupil. This is a large increase and will likely be matched by a jump in New Hampshire's adequacy index when 2001 and 2002 data become available.

Another indicator that did not change over the years 1994 through 2000 is New Hampshire's coefficient of variation. This indicator remained fairly steady and fairly high throughout this time period, meaning that the state still has lot of variation in spending across districts; in 1999, the coefficient was 18.2, the highest of all the years of data. New Hampshire, according to these indexes, still has a great deal of inequity across the districts in the state.

Table 2. Expectations and results by each indicator for New Hampshire

Indicator	Expectations	Results from 1999 to 2000
State equalization effort	↑	↑
Targeting score	←→	↓
Wealth-neutrality score	↑ slightly	↑ slightly
McLoone Index	←→	←→
Coefficient of variation	←→	←→
Adequacy index	↑	↑ slightly
Education spending	↑	↑ slightly
Percent at or above U.S. average	↑	↑

NOTE: ↑ = improved; ↓ = worse; ⟷ = stable

SOURCE: Bureau of the Census: Public Elementary/Secondary Education Finance Data, 1994–2000; Small Area Income and Poverty Estimates, School District Estimates, 1995 and 1997; and School District Data Book, 1990. National Center for Education Statistics: Common Core of Data: Public Elementary/Secondary School District Universe Data, 1994–2000, and *Early Estimates of Public Elementary and Secondary School Education Statistics,* school year 1996–97 to school year 2001–02; and Chambers Cost of Education Index, 1993–94.

Table 3. Changes in equity and adequacy indicators over time in New Jersey

	1994	1995	1996	1997	1998	1999	2000	2001	2002
State equalization effort	47.3	39.8	43.4	43.6	42.5	55.1	48.4	—	—
Targeting score	–.098	–.134	–.125	–.126	–.155	–.170	–.171	—	—
Wealth-neutrality score	.112	.132	.089	.085	.092	.098	.046	—	—
McLoone Index	.892	.894	.906	.911	.904	.946	.916	—	—
Coefficient of variation	—	12.8	11.8	11.5	—	11.7	13.2	—	—
Education spending per student (adjusted for regional cost differences)	—	—	$8,118	$8,176	$8,436	$8,801	$8,667	$9,362	$8,328
Adequacy index	99.97	99.94	99.94	—	—	100	99.99	—	—
Percent of students in districts with per pupil expenditures at or above U.S. average	99.80	99.15	99.15	—	—	99.92	99.78	—	—

—Not available.

SOURCE: Bureau of the Census: Public Elementary/Secondary Education Finance Data, 1994–2000; Small Area Income and Poverty Estimates, School District Estimates, 1995 and 1997; and School District Data Book, 1990. National Center for Education Statistics: Common Core of Data: Public Elementary/Secondary School District Universe Data, 1994–2000, and *Early Estimates of Public Elementary and Secondary School Education Statistics,* school year 1996–97 to school year 2001–02; and Chambers Cost of Education Index, 1993–94.

New Jersey

According to equity and adequacy indicators calculated for New Jersey, there was a noticeable spike in the figures in 1999 (tables 3 and 4). New Jersey showed improvement for that year in its targeting score, state equalization effort, McLoone Index, and education spending per student. The state had a lower targeting score, and a jump in its state share of funding. This combination led to an increase in its state equalization effort from 42.5 in 1998 to 55.1 in 1999. This figure fell back down again slightly in 2000 to 48.4, but still remained an improvement over previous years. The state's McLoone Index also rose in 1999 to .946 from .904, indicating that New Jersey was closer to having all of its students in districts spending at least the median expenditure. Like the state equalization effort, the McLoone Index fell again the next year, to .916, but was still an improvement over previous years.

New Jersey has a similar pattern in its education spending per pupil, only the jump to higher spending did not occur until 2001. The state spent more on its students in the year 2001 than in any other year of these data. The state made the jump from $8,667 in 2000 to $9,362 in 2001. Like the state equalization effort and McLoone Index, this indicator fell the next year back to $8,328.

Table 4. Expectations and results by each indicator for New Jersey

Indicator	Expectations	Results from 1998 to 1999
State equalization effort	↑	↑
Targeting score	↑	↑
Wealth-neutrality score	↑	↑ slightly
McLoone Index	↑	↑
Coefficient of variation	↑	←→
Adequacy index	↑ slightly	↑ slightly
Education spending	↑ slightly	↑ slightly
Percent at or above U.S. average	↑ slightly	↑ slightly

NOTE: ↑ = improved; ↓ = worse; ←→ = stable.

SOURCE: Bureau of the Census: Public Elementary/Secondary Education Finance Data, 1994–2000; Small Area Income and Poverty Estimates, School District Estimates, 1995 and 1997; and School District Data Book, 1990. National Center for Education Statistics: Common Core of Data: Public Elementary/Secondary School District Universe Data, 1994–2000, and *Early Estimates of Public Elementary and Secondary School Education Statistics,* school year 1996–97 to school year 2001–02; and Chambers Cost of Education Index, 1993–94.

From 1994 to 2000, New Jersey has had steady improvement in its wealth-neutrality score. This improvement seems to be even more apparent in 1998 and 1999. New Jersey's wealth-neutrality score was .092 in 1998 and .098 in 1999. These scores indicate that for these years, when both state and local funding are

Table 5. Changes in equity and adequacy indicators over time in Vermont

	1994	1995	1996	1997	1998	1999	2000	2001	2002
State equalization effort	36.4	35.1	34.2	34.8	35.0	87.9	92.8	—	—
Targeting score	–.383	–.364	–.449	–.450	–.401	–.455	–.530	—	—
Wealth-neutrality score	.161	.152	.211	.162	.182	.334	.311	—	—
McLoone Index	.903	.892	.863	.860	.889	.866	.867	—	—
Coefficient of variation	19.0	16.1	16.2	18.6	—	19.2	19.9	—	—
Education spending per student (adjusted for regional cost differences)	—	—	$6,259	$6,764	$6,512	$6,746	$7,408	$8,622	$9,907
Adequacy index	95.49	93.34	92.82	—	—	97.01	97.52	—	—
Percent of students in districts with per pupil expenditures at or above U.S. average	53.83	48.25	48.33	—	—	68.9	81.57	—	—

—Not available.

SOURCE: Bureau of the Census: Public Elementary/Secondary Education Finance Data, 1994–2000; Small Area Income and Poverty Estimates, School District Estimates, 1995 and 1997; and School District Data Book, 1990. National Center for Education Statistics: Common Core of Data: Public Elementary/Secondary School District Universe Data, 1994–2000, and *Early Estimates of Public Elementary and Secondary School Education Statistics,* school year 1996–97 to school year 2001–02; and Chambers Cost of Education Index, 1993–94.

considered, poor and wealthy districts in New Jersey spent similar amounts of money on education.

There was not a great deal of fluctuation in New Jersey's adequacy indicators, however this was expected to some extent since the state's education spending per pupil has consistently been much higher than the national average. The state's adequacy index and the percent of students in districts with per pupil expenditures at or above the national average remained above 99 from 1994 to 2000. The adequacy index reached 100 in 1999, but like the equity indicators, fell slightly in 2000.

Vermont

Vermont had a drastic improvement in its equity grade from *Quality Counts 2001* to *Quality Counts 2002*. This reflected 1997 and 1999 data, respectively. Over this time Vermont had strong changes in two indicators, state equalization effort, which makes up 50 percent of the equity grade, and wealth-neutrality score (tables 5 and 6). Vermont's state equalization effort rose substantially. It was 35.0 in 1998, 87.9 in 1999, and 92.8 in 2000. This is not only a strong change for a state in one year, but it also made Vermont a state with one of the highest state equalization efforts. Vermont has also shown improvement in its targeting score over these years, most recently having a score of –.530 in 2000, the best score of the 50 states.

Table 6. Expectations and results by each indicator for Vermont

Indicator	Expectations	Results from 1998 to 1999
State equalization effort	↑	↑
Targeting score	↑	↑
Wealth-neutrality score	↑ slightly	↓
McLoone Index	↑	↓ slightly
Coefficient of variation	←→	↓ slightly
Adequacy index	↑ slightly	↑ slightly
Education spending	↑ slightly	↑ slightly
Percent at or above U.S. average	↑ slightly	↑ slightly

NOTE: ↑ = improved; ↓ = worse; ←→ = stable.

SOURCE: Bureau of the Census: Public Elementary/Secondary Education Finance Data, 1994–2000; Small Area Income and Poverty Estimates, School District Estimates, 1995 and 1997; and School District Data Book, 1990. National Center for Education Statistics: Common Core of Data: Public Elementary/Secondary School District Universe Data, 1994–2000, and *Early Estimates of Public Elementary and Secondary School Education Statistics,* school year 1996–97 to school year 2001–02; and Chambers Cost of Education Index, 1993–94.

Unfortunately, although Vermont made gains in state equalization effort, it had a worse wealth-neutrality score for these same years. Vermont always had a positive wealth-neutrality score, meaning that property-poor districts, on average, had less state funding per weighted pupil than wealthy districts; however, in recent years its wealth-neutrality score has gotten even

Table 7. Changes in equity and adequacy indicators over time in Wyoming

	1994	1995	1996	1997	1998	1999	2000	2001	2002
State equalization effort	51.5	49.0	56.0	56.9	54.5	56.6	58.1	—	—
Targeting score	.105	.088	–.026	–.093	–.034	.000	–.025	—	—
Wealth-neutrality score	–.153	–.151	–.123	–.202	–.203	–.152	–.189	—	—
McLoone Index	.848	.874	.948	.932	.842	.934	.958	—	—
Coefficient of variation	15.1	13.6	14.7	15.7	—	13.0	12.9	—	—
Education spending per student (adjusted for regional cost differences)	—	—	$6,499	$6,297	$6,590	$6,790	$7,853	$8,657	$8,957
Adequacy index	96.32	96.53	94.47	—	—	100	100	—	—
Percent of students in districts with per pupil expenditures at or above U.S. average	42.92	43.20	37.92	—	—	100	100	—	—

—Not available.

SOURCE: Bureau of the Census: Public Elementary/Secondary Education Finance Data, 1994–2000; Small Area Income and Poverty Estimates, School District Estimates, 1995 and 1997; and School District Data Book, 1990. National Center for Education Statistics: Common Core of Data: Public Elementary/Secondary School District Universe Data, 1994–2000, and *Early Estimates of Public Elementary and Secondary School Education Statistics,* school year 1996–97 to school year 2001–02; and Chambers Cost of Education Index, 1993–94.

higher. Vermont's score rose from .162 in 1997 to .311 in 2000, and peaked at .334 in 1999. This shows that although Vermont has made efforts to increase equity in the state, total state and local funding is still linked to the property wealth of the districts.

Vermont increased its education spending per pupil by more than $3,000 from 1999 to 2002. Spending increased from $6,746 per student in 1999 to $9,907 in 2002. Another adequacy indicator that showed gains over this time was the percent of students in districts with per pupil expenditures at or above the national average. This indicator rose to 81.6 in 2000 from 68.9 in 1999.

Wyoming

Throughout the mid-nineties, Wyoming has shown improvement on three equity indicators: targeting score, state equalization effort, and McLoone Index. In 1996, Wyoming's targeting score became negative, meaning that it started to target funds to property-poor districts. This trend continued for the rest of the years of data analyzed, with the exception of 1999, where Wyoming's targeting score was zero (tables 7 and 8). The state also increased its state equalization effort from 49 in 1995 to 56 percent in 1996. The state's McLoone Index also rose in 1996, from .874 in 1995 to .948, showing that a greater number of students were in districts with expenditures close to the state median.

Table 8. Expectations and results by each indicator for Wyoming

Indicator	Expectations	Results from 1998 to 1999
State equalization effort	↑	↑ slightly
Targeting score	↑	↑ slightly
Wealth-neutrality score	↑	↑
McLoone Index	↑	←→
Coefficient of variation	↑	←→
Adequacy index	↑ slightly	←→
Education spending	↑ slightly	↑
Percent at or above U.S. average	↑ slightly	←→

NOTE: ↑ = improved; ↓ = worse; ←→ = stable.

SOURCE: Bureau of the Census: Public Elementary/Secondary Education Finance Data, 1994–2000; Small Area Income and Poverty Estimates, School District Estimates, 1995 and 1997; and School District Data Book, 1990. National Center for Education Statistics: Common Core of Data: Public Elementary/Secondary School District Universe Data, 1994–2000, and *Early Estimates of Public Elementary and Secondary School Education Statistics,* school year 1996–97 to school year 2001–02; and Chambers Cost of Education Index, 1993–94.

Wyoming's wealth-neutrality score has always been negative, meaning that when local and state funding is considered, property-poor districts, on average, spend more per student than wealthier districts. Most recently, with 2000 data, Wyoming had the best wealth-neutrality score of the 50 states. This indicator has been consistently good with only some fluctuation since 1994.

Wyoming also had a strong rise in funding per student in recent years, from $6,790 in 1999 to $7,853 in 2000. This rise in funding continued for the next 2 years, reaching $8,657 in 2001 and $8,957 in 2002. This increase in funding had a strong effect on the percent of students in districts with per pupil expenditures at or above the U.S. average. In 1996, only 37.9 percent of students were in districts spending at or above the U.S. average, while in 1999 and 2000, 100 percent of students fell into this category.

Discussion of Results

For the most part, the equity and adequacy indicators used in this analysis correspond to the school funding changes documented for each of the four states. Reforms in New Jersey and New Hampshire were very well matched to the indicators, and Vermont and Wyoming also lined up fairly well. There were only two cases (in Vermont) where the results were the opposite of what was anticipated. Twenty-three of a possible 32 indicators (8 indicators by 4 states) matched what was expected from the reforms occurring in the states.

The equity and adequacy indicators used in this analysis correspond to school funding changes documented for each of the states.

In New Hampshire, all indicators matched what was expected with the exception of the targeting score, which was actually worse in 2000. In New Hampshire it was expected that the equity picture would not change very much, since the distribution of funding in the state remained fairly constant even after the new finance system was implemented. The only exception to this was that the state equalization effort was expected to improve with New Hampshire making a greater investment in education. In fact, the state share of funding did increase a great deal in 1999–2000, which was reflected in a strong change in the state equalization effort. For adequacy, it was expected that there should be much more change than equity. The adequacy index and education spending per pupil were expected to go up due to new funding across the board in New Hampshire. Although the adequacy index only rose slightly in 1999 and 2000, education spending per pupil rose substantially between 2000 and 2002. This makes it likely that when data are available for district level calculations based on 2001 and 2002, the adequacy index for those years will be even higher.

For New Jersey, it was expected that adequacy should improve slightly, but that the real gains would be in a clear increase in the equity indicators after 1998. This analysis found that for New Jersey, all indicators except coefficient of variation matched what was expected in light of the reforms implemented by the state. In 1999, New Jersey in fact had a very noticeable spike in its equity indicators, especially its targeting score, state equalization effort, and McLoone Index. The state also showed improvement in its education spending per student, although not until 2001. There was not a lot of change in New Jersey's adequacy indicators, however the state has always done fairly well in this area. An interesting pattern that emerged in the indicators for New Jersey is that after this spike in 1999, most of the indicators fell, which was not expected. This result may be tied to the continuing battle in the state over funding and equity issues.

For Vermont, five of the eight indicators in this analysis matched what was expected from changes made in the state school funding system. Wealth-neutrality score, McLoone Index, and coefficient of variation did not follow the results that were expected. In Vermont, the greatest change was in the state equalization effort for 1999 and 2000, which matches the expectation that the state share of funding would increase after 1998. The targeting score for Vermont also had a fairly strong improvement in 1999 and 2000. For adequacy, it was expected that all three indicators would improve slightly from 1998 to 1999, which matched exactly with the results. In addition, in 2001 and 2002, education spending per pupil in Vermont began to increase rapidly, indicating that in future years of data the adequacy index for the state should continue to improve.

Wyoming was the state in this study with the least congruence between expected and actual results; only four of the eight indicators matched expectations. Most of the changes in equity indicators appeared in 1996, but reform was not implemented until 1998 and 1999. Interestingly, the court made its decision in 1995, so the indicator changes in 1996 may have been a result of smaller adjustments made in how schools are financed inspired by the ruling and implemented before the legislature revised the entire system. One way Wyoming did match expectations was in an increase in education spending per student. Also, it was expected that the adequacy index for the state should improve, and the data show that in 1999 and 2000 Wyoming had a perfect adequacy index. Wyoming also greatly increased its percent of students in districts with per pupil expenditures at or above the U.S. average over this time.

Conclusion

This analysis found that indicators for New Hampshire and New Jersey matched very well with expectations from changes these states made in their school finance systems, and Vermont and Wyoming's results matched fairly well.[6] Even though there was a strong match between indicators and expectations, some indicators were more accurate than others. According to the results for the four states selected for this analysis, the state equalization effort (state share of funding and targeting score) and the three adequacy indicators were well matched to the reforms occurring in the states. The other equity indicators, wealth-neutrality score, McLoone Index, and coefficient of variation were not as clearly matched in all cases, and were less predictable. This reaffirms the weighting system used for grading equity in *Quality Counts*, since state equalization effort constitutes half of the grade, and the other three indicators together constitutes the other half.

It is encouraging to see that for the most part the indicators used in *Quality Counts* reflect the court-mandated changes that occurred in these states. This shows that it is possible to develop accurate assumptions about what these states were doing with their education finance systems based on these indices and the data they are derived from. Interestingly, although it is reassuring to see that these indicators are in fact reflecting true changes in state policy, it is important to have the context of what has happened or is happening in a state when making assumptions based on these numbers.

Another interesting factor evident from this analysis is the problem of the time lag in the availability of federal school finance data because of the difficulty in collecting and standardizing data across all 50 states and the District of Columbia. For *Quality Counts 2003*, the most recent data available for these indicators reflected the 1999–2000 school year. During the time between when these data are collected and when they are published, the states could have implemented drastic changes to their school finance systems. This is another reason that contextual information—especially current information—is important to consider when making assumptions about these indicators and how they are changing over time.

In part due to the results of this study, *Education Week* is conducting a state policy survey on how states raise revenues and distribute funds for education. Some of these data will be published in *Quality Counts 2004*, and more will be included in a regular issue of *Education Week* in the winter of 2004. These data will serve to not only help inform the indicators described in this paper and used in *Quality Counts*, but also will be a tool for school finance researchers to use as current background information and context for their analyses.

[6] One reason Vermont and Wyoming may have had less clear results from this analysis is that reforms in these states were implemented over several years, and this analysis looked for changes in indicators in the single year where the most changes occurred.

Appendix

After the data files were downloaded and merged for the most recent year available, the next step was to eliminate districts that met certain characteristics. Districts with certain characteristics were eliminated because the purpose of this analysis was to measure equity and adequacy in public elementary and secondary schools only. Districts were deleted if they met the following conditions: school levels other than elementary, secondary, or unified (SCHLEV should be 1, 2, or 3 only), no schools (SCH## = 0), state or federal level (TYPE## = 3, 4, 5, 6, or 7), and fall membership (V33) less than 200.

After eliminating appropriate data, the next step was to calculate the variables needed for the indicators. Following is a list of all the equations and calculations that were used in this study.

1. Adjusted State Revenue per Pupil Index
   a. Total State Enrollment (TSE) = (Σ V33 for each state)
   b. Share of Total State Enrollment (STSE)=V33 / TSE
   c. State-Indexed Cost of Education Index (SICEI) = CEIL93 / (Σ (STSE * CEIL93) for each state)
   d. Adjusted Cost of Education Index (ADJCEI) = (0.85 * SICEI) + 0.15 (Assign a 1 if missing data)
   e. Adjusted State Revenue (ADJSTRV) = (TSTREV / ADJCEI)
   f. Adjusted State Revenue per Pupil (ADSTRVPP) = (ADJSTRV * 1000) / V33
   g. Total Adjusted State Revenue for Each State (TADSTREV) = (Σ ADJSTRV for each state)
   h. Average Adjusted State Revenue per Pupil for Each State (AVGASTPP) = (TADSTREV / TSE) * 1000
   i. Adjusted State Revenue per Pupil Index (ADSPPIND) = (ADSTRVPP / AVGASTPP)

2. Adjusted District Wealth per Pupil Index
   a. Adjusted District Wealth (ADJWLTH) = (WEALTH / ADJCEI)
   b. Total Adjusted District Wealth (TADWTH) = (Σ ADJWLTH for each state)
   c. Adjusted District Wealth per Pupil (ADJDWPP) = (ADJWLTH / V33)
   d. Average Adjusted District Wealth per Pupil (AVGDWPP) = (TADWTH / TSE)
   e. Adjusted District Wealth per Pupil Index (ADDWIND) = (ADJDWPP / AVGDWPP)

3. Percent of Students in Poverty Index
   a. Estimated Percentage of Children 5 to 17 in Poverty (POVPER) = (POORCHRN / CHILD)
   b. Estimated Number of Children in Poverty (NUMPOV) = POVPER * V33
   c. Total Estimated Number of Children in Poverty for Each State (TPOVST) = (Σ NUMPOV for each state)
   d. Average Percentage of Children in Poverty for Each State (POVAVG) = (TPOVST / TSE)
   e. Percent of Students in Poverty Index (POVINDEX) = (POVPER / POVAVG)

4. Percent of Special Education Students Index
   a. Total Number of Special Education Students in Each State = (Σ SPECED## for each state)
   b. Percent of Enrollment that is Special Education (PERIEP) = (SPECED## / V33)
   c. Average Percent Enrollment for Each State (IEPAVG) = (TOTSPCED / TSE)
   d. Percent Special Education Students Index (IEPINDEX) = (PERIEP / IEPAVG)

5. Enrollment-Squared Index
   a. District Enrollment Squared (V33_2) = $(V33)^2$
   b. Total District Enrollment Squared for Each State (TOTV33_2) = (Σ V33_2 for each state)
   c. District Enrollment Squared per Pupil Index (ENSQUIND) = (V33_2 / TOTV33_2) / (V33 / TSE)

6. Land Area per Pupil Index
   a. Area per Pupil (AREAPP) = (AREA / V33)
   b. Total Area per State (TSAREA) = (Σ AREA for each state)
   c. Average Area per Pupil for Each State (AVGARPP) = (TSAREA / TSE)
   d. Area Per Pupil Index (AREAINDX) = (AREAPP / AVGARPP)

7. Weighting Variable (WGT) = (V33 / TSE)* Number of Districts in Each State

8. State Share of Funding
   a. Adjusted Local Revenue (ADJLOCRV) = (LOCREV / ADJCEI)

b. Adjusted State & Local Revenue (STLOCREV) = (ADJLOCRV + ADJSTRV)
c. Total Adjusted State & Local Revenue (TASLR) = (Σ STLOCREV for each state)
d. State Share of Funding for Each State (STSHARE) = (TADSTREV / TASLR)

9. Wealth-Neutrality Score
   a. Weighted Enrollment (WGTENROL) = (SPECED## * 1.3) + (NUMPOV * 0.2) + V33
   b. Total Weighted Enrollment for Each State (TOTWTENR) = (Σ WGTENROL for each state)
   c. Adjusted Local Revenue (ADJLOCRV) = (LOCREV / ADJCEI)
   d. Adjusted State & Local Revenue (STLOCREV) = (ADJLOCRV + ADJSTRV)
   e. Adjusted State & Local Revenue per Weighted Pupil (ASLRPWP) = (STLOCREV / WGTENROL)
   f. Total Adjusted State & Local Revenue (TASLR) = (Σ STLOCREV for each state)
   g. Average Adjusted State & Local Revenue per Weighted Pupil (AVASLRWP) = (TASLR / TOTWTENR)
   h. Adjusted State & Local Revenue per Weighted Pupil Index (ASLRPIND) = (ASLRPWP / AVASLRWP)
   i. Adjusted District Wealth per Weighted Pupil (ADWWP) = (ADJWLTH / WGTENROL)
   j. Average Adjusted District Wealth per Weighted Pupil (AADWPWP) = (TADWTH / TOTWTENR)
   k. Adjusted District Wealth per Weighted Pupil Index (ADJWWP) = (ADWWP / AADWPWP)

10. McLoone Index
   a. Adjusted Expenditures (ADJEXP) = (TCURELSC / ADJCEI)
   b. Adjusted Expenditures per Weighted Pupil (ADJPWP) = (ADJEXP * 1000) / WGTENROL

11. Adequacy Index
   a. Adjusted Cost of Education Index for Adequacy Indicators (ADJCEIAD) = (State level CEI for districts missing CEI data, else CEI)
   b. Adjusted Expenditures for Adequacy Indicators (ADJEXPAQ) = (TCURELSC / ADJCEIAQ)
   c. Adjusted Expenditures per Weighted Pupil for Adequacy Indicators (ADJPWPAQ) = (ADJEXPAQ * 1000) / WGTENROL

References

Advocacy Center for Children's Educational Success With Standards. (n.d.). *Finance Litigation: New Jersey.* Retrieved March 26, 2003 from http://www.accessednetwork.org/litigation/lit_nj.html

Claremont School District v. Governor, 635 A.2d 1375 (N.H. 1993). (Claremont I)

Claremont School District v. Governor, 703 A.2d 1353 (N.H. 1997). (Claremont II)

Claremont School District v. Governor, 744 A.2d 1107 (N.H. 1999). (Claremont III)

Education Week (2001, January 11) *Quality Counts 2001: A Better Balance.*

Education Week (2002, January 10) *Quality Counts 2002: Building Blocks for Success*.

Education Week (2003, January 9) *Quality Counts 2003: "If I Can't Learn From You."*

Gold, S., Smith, D., and Lawton, S. (Eds.). (1995). *Public School Finance Programs of the United States and Canada, 1993–94*. Albany, NY: The American Education Finance Association and The Nelson A. Rockefeller Institute of Government.

Guthrie, J.W., Hayward, G.C., Smith, J.R., Rothstein, R., Bennett, R.W., Koppich, J.E., Bowman, E., DeLapp, L., Brandes, B., and Clark, S. (1997). *A Proposed Cost-Based Block Grant Model for Wyoming School Finance.* Retrieved March 23, 2003, from http://legisweb.state.wy.us/school/cost/apr7/apr7.htm

Hall, D.E. (2002). *School Finance Reform: The First Two Years* (Plumbing the Numbers #6). Concord, NH: New Hampshire Center for Public Policy Studies.

Heaps, R., and Woolf, A. (1997, October). *Vermont's New Education Financing Law: How it Works, What it Means.* Retrieved March 23, 2003, from http://www.vtroundtable.org/siteimagesvtsnewedfinancinglaw.pdf

Jimerson, L. (2002). *Still "A Reasonably Equal Share": Update on Educational Equity in Vermont Year 2001–2002.* Retrieved March 23, 2003, from http://ruraledu.org/docs/vtequity/vt_rep02_72ppi.pdf

Mathis, W.J. (2000, April). *Vermont's Act 60: Early Effects of a Comprehensive School Finance Reform*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.

Miller, L. (1995, November 22). Court Again Strikes Down Wyo. Finance System. *Education Week,* Retrieved March 23, 2003, from http://edweek.org/ew/ew_printstory.cfm?slug=12wyo.h15

National Center for Education Statistics. (2001). *Public School Finance Programs of the United States and Canada: 1998–99* (NCES 2001–309). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Newman, M. (1990, June 13). Finance System for N.J. Schools Is Struck Down. *Education Week*, Retrieved March 26, 2003, from http://www.edweek.org/ew/ewstory.cfm?slug=09440001.h09

Robinson v. Cahill, 303 A.2d 273 (N.J. 1973).

Vt. Stat. Ann. tit. 16, § 4011 (2003).

Viadero, D. (2001, July 11). N.H. Lawmakers OK Finance Plan, But Debate Lives On. *Education Week*, Retrieved March 16, 2003, from http://edweek.org/ew/ew_printstory.cfm?slug=42nh.h20

School Finance Reform and School Quality: Lessons From Vermont

Thomas Downes

Department of Economics
Tufts University

About the Author

Thomas A. Downes is an associate professor of economics at Tufts University. He received his B.A. in economics and mathematics from Bowdoin College in 1982 and his Ph.D. from Stanford University in 1988 and was a faculty member at Northwestern University from 1988–1994. Professor Downes' research focuses primarily on the evaluation and construction of state and local policies to improve the delivery of publicly provided goods and reduce inequities in these services, with particular attention paid to public education. He has also pursued research on the roles of the public and private sectors in the provision of education. Among his current projects are papers that document the implications for student outcomes and community composition of school finance reforms and property tax limitations and that quantify the extent to which charter schools draw their enrollments from private schools. His published papers have appeared in such journals as *Public Finance Quarterly,* the *National Tax Journal, Public Choice,* the *RAND Journal of Economics, Bulletin of Economic Research, Economics of Education Review,* and the *Journal of Urban Economics.* Professor Downes served as a member of the Panel on Formula Allocations of the Committee on National Statistics of the National Academies, was a participant in the symposium sponsored by the New York State Board of Regents that resulted in the Board's report, *Educational Finance to Support High Learning Standards,* and currently is a member of the board of the American Education Finance Association. He can be contacted at thomas.downes@tufts.edu.

School Finance Reform and School Quality: Lessons From Vermont

Thomas Downes

Department of Economics
Tufts University

Acknowledgments

Thanks to Tom Husted, Bill Mathis, Larry Kenny, Jim Wyckoff, John Yinger, Bob Strauss, Bruce Baker, and participants in the Economics Department seminar at the University of New Hampshire, in the Maxwell School's Conference on State Aid to Education, and in the 2001 meetings of the American Education Finance Association and the Association for Public Policy Analysis and Management for their helpful comments and suggestions. All errors of omission and commission are the responsibility of the author.

Introduction

In June of 1997, the elected leaders of Vermont enacted the Equal Educational Opportunity Act (Act 60) in response to a state supreme court decision in the *Brigham v. State* case. Act 60 may well have represented the most radical reform of a state's system of public school financing since the post-*Serrano*, post–Proposition 13 changes in California in the late 1970s. As a result, Act 60 could provide a unique opportunity to determine if dramatic school finance reforms like those enacted in Vermont generate greater equality in measured student performance. This paper represents an attempt to document the changes in the distributions of spending and of student performance that have occurred in the post–Act 60 period.

This paper begins with an overview of the institutional structure of educational finance and provision in Vermont. One purpose of this overview is to make the argument that the Vermont case is particularly interesting because there have not been dramatic demographic changes in the state that could obscure the impact of finance reforms. With this context established, I review the research on the links between finance reforms and the distributions of education spending and of student performance. After briefly discussing the data utilized, I examine the extent to which there has been convergence across school districts in expenditures and in student performance.

All of the available data support the conclusion that the link between spending and taxable resources has been significantly weakened and that spending, however it is measured, is substantially more equal. I also present evidence that the cross-district dispersion of the performance of fourth-graders on standardized tests of mathematics has declined post–Act 60. And there is no evidence of increased cross-district dispersion of the test performances of second- and eighth-graders.

School Finance Reform in Vermont

In 1995 in Lamoille Superior Court, a group of plaintiffs that included Amanda Brigham, a 9-year-old student in the Whiting School District (Burkett 1998), filed suit against the State of Vermont. The goal of the suit was to force substantive reform of a system of school financing that the plaintiffs felt deprived students in property-poor school districts of equal educational opportunities and forced taxpayers in these property-poor districts to assume a disproportionate burden of the financing of public education. On February 5, 1997, the Supreme Court of the State of Vermont ruled in favor of the plaintiffs, concluding that the existing system deprived "children of an equal educational opportunity in violation of the Vermont Constitution" (Brigham v. State, 166 Vt. at 249). The court left it to the legislature to craft a new financing system that was consistent with the state constitution.

The Vermont court, concluding the existing system deprived "children of an equal educational opportunity in violation of the Vermont Constitution," left it to the legislature to craft a new financing system.

The focus in the plaintiffs' suit on both inequalities in educational spending and disparities in property tax burdens grew out of longstanding dissatisfaction in Vermont with the existing foundation system of education financing and the existing system of property taxation. Prior to Act 60, Vermont used a traditional foundation formula to determine the state aid a town received:

(1) Total state aid = (Weighted average daily membership) * (Foundation amount) – (Foundation tax rate) * (Aggregate fair market value) * 0.01,

where the weighted average daily membership was determined by assigning weights of 1.25 to secondary students and students receiving food stamps, assigning weights of between 1.0385 and 1.0714 to students who must be transported to school, and averaging the weighted counts from the previous 2 school years (Mathis 1995). While the foundation amount was set with the intent of permitting districts spending that amount to meet state standards for those students assigned a weight of 1, fluctuations in the state's fiscal status led the state legislature to adjust the foundation tax rate so as to reduce the state's aid liability. As a result, the state share of basic educational expenditures fluctuated between 0.20 and 0.37, with the share declining when the state economy weakened (Mathis 2001). The period leading up to Act 60 was a period of decline in the state share.

The widespread dissatisfaction with the existing school financing system had not been ignored by elected officials. In both 1994 and 1995, the state house of the Vermont Legislature approved legislation designed to overhaul education financing. While this legislation failed to pass the state senate, the legislation contained key elements of the eventual response to the *Brigham* decision (McClaughry 2001).

The legislation, by highlighting concerns about education financing and property taxation, also influenced the dynamics of the 1996 election. The state senate that was elected in 1996 was committed to property tax reform (Mathis 2001). The result was a state legislature that was ready to move on legislation that would comply with the *Brigham* decision and reduce the property tax burdens of poor individuals.

Given the political dynamic in Vermont, the speed with which Act 60, the legislation designed to comply with *Brigham* and to provide property tax relief, was passed surprised no one. Signed into law on June 26, 1997, Act 60 created a system of school financing that combined elements of foundation and power equalization plans. A statewide property tax was established, with revenues from the tax being used to finance a portion of foundation aid.[1] If in a locality property tax revenues generated by levying the statewide rate exceed the amount needed to finance the foundation level of spending, the excess property tax revenues are recaptured by the state.

[1] In the 2000–2001 school year, the nominal property tax rate was 1.1 percent, and the foundation level was $5,200.

Under Act 60, localities are allowed to choose spending levels in excess of the foundation level. To weaken the link between property wealth and spending in excess of the foundation level, the act established a power equalization scheme that insured that localities with the same nominal tax rates would have the same levels of education spending. The power equalization scheme also included a unique recapture element; all spending in excess of the foundation level is drawn from a sharing pool that consists of all the property tax revenues generated by property tax rates in excess of the statewide rate. As a result, no revenues from statewide taxes are used to finance the power equalizing portion of the school finance system. Further, when the voters in a locality choose a nominal property tax rate above the statewide rate, the revenues that will be available for that locality's schools will not be known with certainty until all other localities have made their taxing decisions and the size of the sharing pool is established.

While the *Brigham* decision forced state policymakers to implement finance reforms, the reality was that Act 60 was as much about property tax relief as it was about school finance reform. For taxpayers in many communities, the finance reforms by themselves would have dramatically reduced tax burdens by allowing localities to maintain or even increase education spending with substantially lower tax rates. At the same time, taxpayers in high-wealth communities, which have been labeled "gold towns," necessarily faced increases in their property tax payments.[2] To lessen the burden on low-income residents of the "gold towns," the drafters of Act 60 included in the legislation a provision that granted tax adjustments to certain homestead owners. These tax adjustments were explicitly linked to the taxpayer's income; the original legislation specified that all owners with incomes at or below $75,000 were eligible for adjustments.

The reality was that Act 60 was as much about property tax relief as it was about school finance reform.

All of these changes in the property tax were clearly designed to shift some of the burden of financing Vermont's schools away from state residents to corporations and nonresident owners of property in Vermont.[3]

Thus, Act 60 continues the recent tradition of linking school finance reforms and tax relief that is exemplified by Michigan. For this reason, any complete evaluation of the success of Act 60 must consider both the changes in education provision and the changes in tax burdens. Therefore, this paper necessarily provides a partial view of the welfare implications of Act 60.[4]

The school finance reforms that were the central element of Act 60 were phased in over several years, with the new regime not fully in place until the 2000–2001 academic year. Nevertheless, as was true in California in the aftermath of *Serrano* and Proposition 13, in some districts there were surprisingly rapid responses to Act 60. Not surprisingly, in the gold towns there was vocal opposition to Act 60. Also unsurprising, given the California experience, were the efforts in these towns to encourage residents to make voluntary contributions to the schools[5] and to shift to town governments responsibility for financing certain "school" functions.

2 In the 1994–95 school year, 69 of the 248 towns in Vermont for which data were available had effective education property tax rates below $1.10 per $100 in assessed value. While the percentage of towns with effective education rates below $1.10 had undoubtedly declined by the 1997–98 school year, the last year before the phasing in of Act 60 began, the reality was still that Act 60 forced a sizable fraction of the towns in Vermont to increase property tax rates.

3 The correlation between each town's effective education property tax rate in 1994–95 and the fraction of that town's property that was owned by town residents in 1998–99 was 0.5461. In other words, towns with low effective property tax rates prior to Act 60 also tended to be towns in which a large fraction of the property tax burden was exported.

4 See Heaps and Woolf (2000) and Jimerson (2001) for efforts to evaluate both the implications of Act 60 for educational provision and the effects of Act 60 on property tax burdens.

5 Since towns that collected sufficient funds from individual contributions could avoid participating in the sharing pool, in most gold towns education funds were established and property owners were encouraged to contribute to these funds. Participation rates varied across towns, ranging up to 87 percent in Manchester, where aggressive tactics, such as publication of the names of nonparticipants, were used to encourage giving. More traditional incentives were also used to encourage giving; in fiscal years 1999, 2000, and 2001, the Freeman Foundation matched individual donations to the funds.

Another California parallel is the apparent growth in private school enrollments that have been mentioned in press reports.

Care needs to be taken, however, to avoid making too much of the California parallels. Act 60 gave Vermont school districts much more discretion over the level of expenditures than California districts have. The tax price of education spending has increased in the gold towns, but spending is not being forcibly leveled down as it was in California. Also, low-wealth towns were not required to maintain local effort; several towns used the Act 60 windfall primarily to reduce nominal property tax rates. As a result, low-wealth towns were not necessarily leveled up. The reality is that Act 60 did not duplicate the California reforms; a fact on which the next section expands.

Act 60 did not duplicate the California reforms in one other important way. The California reforms predated the nationwide push for accountability; Act 60 was passed at a time when most states were attempting to strengthen accountability and educational standards. As a result, several elements of the legislation built on the existing system of testing and standards to strengthen accountability. For example, under Act 60 all districts were required to develop action plans to improve student performance on the tests that are part of the Vermont Comprehensive Assessment System. In addition, the state board of education was mandated to take on a more active oversight role. Nevertheless, the central elements of the state's accountability system were unaffected by Act 60.

Act 60 gave Vermont school districts much more discretion over the level of expenditures than California districts have.

Why Study Vermont?—A Review of Research on the Impact of School Finance Reforms

The school finance reforms implemented in California in the aftermath of that state's supreme court decision in the *Serrano v. Priest* case and of the nearly contemporaneous tax limits imposed by Proposition 13 represent a watershed both in the debate over the structure of school finance reforms and in the direction of research into the impact of those reforms. In the post-*Serrano* period, the California reforms and their supposed effects on the schools in that state have been discussed in every state in which school finance reforms have been implemented. Vermont is no exception; the supposed parallels between the California reforms and Act 60 have been mentioned repeatedly.[6]

The California reforms also shifted the focus of research on the impact of school finance. Prior to the reforms, the focus in the literature was almost solely on the impact of finance reforms on spending inequality. After *Serrano*, the scope of the analysis broadened to include the impact of finance reforms on the level and distribution of student achievement, on housing prices, on the supply of private schooling, and even on the composition of affected communities.[7] The California reforms also became the touchstone for theoretical work. Papers like those of Nechyba (1996, 2000), Bènabou (1996), and Fernandez and Rogerson (1997, 1998) used a California-like system as the post-reform case when trying to reach predictions about the likely effects of finance reform.

The problem with using the California case as a benchmark is that the case has proven to be the exception, not the rule. First, the limits imposed on local control over spending have not been duplicated in any other state. Even in Michigan and Vermont, the states in which the most extensive post-*Serrano* reforms have been implemented, some degree of local control over taxes and spending is permitted. Further, the population of students served by California schools changed more dramatically than the population of students in any other state in the nation. From 1986 to 1997, the percentage of the California public school student

6 For examples of references to California, see McClaughry (1997) and Mathis (1998).

7 The number of papers dealing with these varied topics are too numerous too cite. Evans, Murray, and Schwab (1999) and Downes and Figlio (1999, 2000) cite many of the relevant papers.

population identified as minority increased from 46.3 percent to 61.2 percent. Nationally, the percent minority grew far more slowly, from 29.6 percent to 36.5 percent.[8] As Downes (1992) notes, these demographic changes make it difficult to quantify the impact of the finance reforms in California on the cross-district inequality in student achievement.

The possibility that California might be the exception and not the rule pushed a number of researchers to pursue national-level studies attempting to document the impact of finance reforms. On the spending side, Silva and Sonstelie (1995), Downes and Shah (1995), and Manwaring and Sheffrin (1997) all took slightly different approaches to quantifying the effect of finance reforms on mean per pupil spending in a state. Because they used district-level data, Hoxby (2001), Evans, Murray, and Schwab (1997), and Murray, Evans, and Schwab (1998) were able to consider not only the effects of finance reforms on mean spending but also the extent to which spending inequities were reduced by those reforms. As a result, these studies provide the most obvious sources for predictions of the long-run effects of Act 60. The problem is that these studies generate contradictory predictions. Hoxby's results would lead us to expect leveling down, since Act 60 dramatically increases tax prices in towns with more property wealth. Murray, Evans, and Schwab conclude that court-mandated reforms like Act 60 typically result in leveling up.

The same lack of a clear prediction would be apparent to the reader of national-level attempts to determine how the distribution of student performance in a state is affected by a finance reform. Hoxby (2001) represents the first attempt to use national-level data to examine the effects of finance reforms on student performance. She finds that dropout rates increase about 8 percent, on average, in states that adopt state-level financing of the public schools. Although Hoxby's work does not explicitly address the effect of equalization on the within-state distribution of student performance, it seems likely that much of the growth in dropout rates occurred in those districts with relatively high dropout rates prior to equalization. In other words, these results imply that equalization could adversely affect both the level and the distribution of student performance.

Dropout rates increase about 8 percent, on average, in states that adopt state-level financing of the public schools.

While the dropout rate is an outcome measure of considerable interest, analyses of the quality of public education in the U.S. tend to focus on standardized test scores and other measures of student performance that provide some indication of how the general student population is faring. Husted and Kenny (2000) suggest that equalization may detrimentally affect student achievement. Using data on 34 states from 1976–77 to 1992–93, they find that the mean SAT score is higher for those states with greater within-state spending variation. However, the period for which they have test score information, 1987–88 to 1992–93, postdates the imposition of the first wave of finance reforms. Thus, the data do not permit direct examination of the effects of policy changes. In addition, because they use state-level data, Husted and Kenny cannot examine the degree to which equalization affects cross-district variation in test scores.[9] Finally, since only a select group of students take the SAT, Husted and Kenny are not able to consider how equalization affects the performance of all students in a state.

Card and Payne (2002) explore the effects of school finance equalizations on the within-state distributions

8 Generating comparable numbers for earlier years is difficult. Nevertheless, the best available data support the conclusion that these sharp differences in trends in the minority share predate the *Serrano*-inspired reforms. For example, calculations based on published information for California indicate the percentage minority in 1977–78 was approximately 36.6 percent. Nationally, estimates based on the October 1977 Current Population Survey indicate the percent minority was 23.9 percent.

9 Husted and Kenny do find evidence consistent with the conclusion that, in states which have school finance reforms, these reforms have no impact on the standard deviation of SAT scores. Since, however, the standard deviation of test scores could be unchanged even if cross-district inequality in performance had declined, this evidence fails to establish that finance reforms do not reduce cross-district performance inequality.

of SAT scores. They characterize a school finance policy as more equalizing the more negative the within-state relationship between state aid to a school district and school district income is. They find that the SAT scores of students with poorly educated parents (their proxy for low income) increase in states that, under their definition, become more equalized. Data limitations, however, make it impossible for Card and Payne to examine the effects of policy changes on students residing in school districts in which the changes had the greatest impact. Moreover, while Card and Payne correct for differences in the fractions of the population taking the SAT test, it is still very likely that the students who come from low-education backgrounds but take the SAT test are a very select group and are extremely unlikely to be representative of the low-income or low-education population as a whole.[10]

Downes and Figlio (2000) attempt to determine how the tax limits and finance reforms of the late 1970s and early 1980s affected the distribution of student performance in states in which limits were imposed and how student performance has changed in these states relative to student performance in states in which no limits or finance reforms were imposed. The core data used in the analysis were drawn from two national data sets, the National Longitudinal Study of the High School Class of 1972 (NLS:72) and the 1992 (senior year) wave of the National Education Longitudinal Study of 1988 (NELS:88/92). The NELS data were collected sufficiently far from the passage of most finance reforms to permit quantification of the long-run effects of these reforms by analyzing changes in the distributions of student performance between the NLS:72 cross-section and the NELS cross-section.

Finance reforms in response to court decisions result in small and frequently insignificant increases in the mean level of student performance.

Downes and Figlio find that finance reforms in response to court decisions, like that in the *Brigham* case, result in small and frequently insignificant increases in the mean level of student performance on standardized tests of reading and mathematics. Further, they note that there is some indication that the post-reform distribution of scores in mathematics may be less equal. This latter result highlights one of the central points of the paper; any evaluation of finance reforms must control for the initial circumstances of affected districts. The simple reality is that finance reforms are likely to have differential effects in initially high-spending and initially low-spending districts.

The fundamental reason for the absence of clear predictions of the impact of finance reforms has been mentioned by a number of authors (e.g., Downes and Shah [1995], Hoxby [2001], Evans, Murray, and Schwab [1997]), all of whom have emphasized the tremendous diversity of school finance reforms. In a national-level study, any attempt to classify finance reforms will be imperfect. So, even though there is general consensus that the key elements of a finance reform are the effects of the reform on local discretion, the effects of the reform on local incentives, and the change in state-level responsibilities in the aftermath of reform (Hoxby [2001], Courant and Loeb [1997]), different authors take different approaches to account for the heterogeneity of the reforms. The result is variation in predictions generated by studies that are asking the same fundamental question. The answer, it seems, is not to try to improve the methods of classifying reforms, but is instead to carefully analyze certain canonical reforms. Act 60 is likely to be just such a canonical reform.

In looking for guidance for an analysis of the Vermont reforms, the first case to consider is that of Kentucky, where the reforms that followed a court decision invalidating the system of school finance may represent the most radical change to a state's system of public schooling provision. Flanagan and Murray (2002) document the effects of the reforms in Kentucky. Unfortunately, because the reforms in Kentucky were so extensive, any lessons from that case are probably not

[10] For instance, among the students in Card and Payne's low-parental-education group, in 28 states in 1978 (25 states in 1990) fewer than 10 percent took the SAT examination and in 20 states in 1978 (15 states in 1990) fewer than 3 percent took the SAT. Further, in 1978 no state had more than 36.2 percent of the low-parental-education group take the SAT.

particularly relevant for those attempting to predict the effect of reforms that, like Act 60, primarily affect the system of school finance.

Thus, the most direct antecedent in this case-study approach to analyzing finance reforms is Downes (1992), who showed that the extensive school finance reforms in California in the late 1970s generated greater equality across school districts in per pupil spending but not greater equality in measured student performance. Duncombe and Johnston's (2002) work on Kansas offers an example of a recent case study of a canonical reform. This study of Vermont is another such example. Will the outcomes in Vermont duplicate those in California? What are the similarities in and differences between the results for Vermont, Kansas, and Kentucky? The data used to answer these questions are described in the next section.

Data

Sources

The majority of the data that are analyzed in this paper are drawn from the Vermont School Report and from publications of the Vermont Department of Taxes. In addition, town-level data on school expenditures were drawn from Heaps and Woolf (2000) and from files created by the Vermont Department of Education and posted at http://www.state.vt.us/educ/new/html/data/perpupil.html. The Vermont Indicators Online database, which is maintained by the Center for Rural Studies at the University of Vermont, was the source of some pre-1999 information on income, demographics, and property wealth at the town level. Finally, the Common Core of Data, maintained by the National Center for Education Statistics, was the source of school-level data on the racial/ethnic composition of each school's student body.

The extensive school finance reforms in California generated greater equality across school districts in per pupil spending but not greater equality in measured student performance.

The norm in Vermont is that towns and school districts are coterminous. There are, however, numerous deviations from the norm. Some small towns do not operate elementary or secondary schools; the children from these towns are sent to public or even private schools in neighboring communities, with tuition payments going from the sending towns to the receiving schools.[11] Many other towns do not have their own high schools, choosing to either "tuition out" their high school students or to participate in unit high school districts.[12] Since one of the goals of this research was to quantify the impact of Act 60 on the inequality in services provided to the schoolchildren of Vermont, the school district had to be the fundamental unit of analysis. Thus, several decisions had to be made to ensure that what was presented was the most accurate picture of the impact of Act 60 on the distributions of expenditures and student performance.

First, all towns that were not tuitioning out students at the elementary level were matched to the school district serving elementary school students from that town. The same matching process was done for towns not tuitioning out high school students.[13] Knowing the town-school district matches made it possible to create school-district-level versions of some variables that were only available at the town level. Second, towns were grouped into types based on institutional arrangements. This made it possible to examine separately the impact of Act 60 on school districts linked to towns with different institutional arrangements.

Nevertheless, the reality in Vermont is that school spending levels are voted on in town meetings, that state aid flows to towns and not school districts, and that analyses of the impact of Act 60 have tended to

[11] In the 2001–02 school year, 824 equalized pupils (out of 103,347 equalized pupils in the state) resided in towns or other areas in which all students were "tuitioned out."Another 87 equalized pupils resided in towns that did not operate an elementary school but belonged to a union high school district.

[12] In the 2001–02 school year, 15,274 equalized pupils resided in towns in which elementary students were served locally but high school students were tuitioned out. Of these 15,274 equalized pupils, less than half were tuitioned out.

[13] If a town tuitions out either elementary or high school students, those students could be attending school in several surrounding districts. As a result, the town cannot be matched to a single elementary or high school district.

focus on variation in expenditures across towns. So, even though cross-town variation provides an imperfect indication of the variation across school districts (and, thus, across students) in expenditures, results are presented that use town-level data on expenditures. These results make it possible to compare the findings in this study to those in previous work. Further, the town-level data are such that it is possible to make a crude adjustment for the effect of institutional variation on expenditures. In particular, the analysis in this paper will use two alternative measures of expenditures, one of which is explicitly designed to adjust for variation in institutional structure.

Summary Statistics

One of the advantages of examining the impact of finance reforms in Vermont is the stability of the student population served by Vermont schools. For example, in the 1995–96 school year 3.12 percent of the students attending public school in Vermont were identified as minority. This percentage fluctuated slightly over the next 4 academic years, from 2.73 percent in 1996–97 to 3.16 percent in 1999–2000.[14] Clearly, in Vermont, unlike in California, the schools were not trying to adjust to a dramatically changing population at the same time they were coping with the effects of finance reforms.

One of the advantages of examining the impact of finance reforms in Vermont is the stability of the student population.

Other measures of the income and demographics of the Vermont student population were also relatively stable both immediately before and just after the implementation of Act 60. For each school year from 1994–95 to 2000–01, table 1 provides summary statistics on certain key measures of the demographics and income of each school district in the state. Average adjusted gross income per exemption, a rough proxy for per capita income, did increase throughout the period, an unsurprising result given that all dollar figures in the table are nominal and that this was a period of strong economic expansion. What is more striking, however, is the stability across time of the poverty rate and the percent of students eligible for free or reduced-price school lunches. The observable characteristics of the population of students being served by Vermont schools appear to have changed little over time.

This stability of measured attributes of the student population does not insure that there have been no significant changes in critical unmeasured characteristics of the students served by the public schools in Vermont. In other words, in an event analysis of the impact of Act 60 on the distribution of student performance, there will be no way to rule out the possibility that cross-time changes in the distribution are driven by cross-time changes in unobservables as opposed to by the effects of the finance reforms. That said, the Vermont context still provides researchers with the best opportunity to date to estimate the effects of finance reforms on the distribution of student performance.

The remaining rows in table 1 provide summary information on some of the expenditure and student performance measures available in the Vermont School Report. No obvious trends in performance are apparent in table 1. Some performance measures improved after Act 60; others declined. For some of the measures of performance, dispersion fell after Act 60, but dispersion increased for other measures. However, because the crude summary measures in table 1 give no indication of how post–Act 60 changes are linked to a district's pre–Act 60 status, no conclusions about the performance effects of Act 60 can be drawn on the basis of the evidence in table 1.

Table 1 also does not support any firm conclusions about the extent to which the link between local wealth and spending has been weakened by Act 60. That said, even the summary measures in table 1 provide some indication of the impact of Act 60 on the dispersion in expenditures. The coefficient of variation of current expenditures per pupil increased from 3.61 in 1994–95 to 5.61 in 2001–02. While some of this increase predated Act 60, Act 60 has mattered, a fact that

[14] Means across schools of the percent minority evidence the same stability. In 1995–96, the across-school mean was 2.1 percent. In 1999–2000, the across-school mean was 2.8 percent.

Table 1. Summary statistics—selected characteristics of Vermont school districts

Variable	Pre–Act 60								Post–Act 60							
	1994–95		1995–96		1996–97		1997–98		1998–99		1999–2000		2000–01		2001–02	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Current expenditures per pupil (in dollars)	5572.94	1523.49	5773.13	1512.51	5935.14	1012.70	6175.11	1105.13	6652.33	1294.00	7601.18	1330.29	8262.36	1491.91	8888.53	1583.64
Special education costs per eligible pupil (in dollars)	1208.36	635.58	1302.94	719.15	1363.33	654.22	1503.53	710.10	1693.59	839.70	—	—	—	—	—	—
Students per teacher	17.96	11.24	25.52	60.16	17.28	3.28	16.86	3.04	16.57	3.56	15.80	2.92	15.18	2.88	15.03	3.41
Average teacher salary (in dollars)	32464.85	4237.02	33530.49	4487.78	33948.62	4495.41	—	—	34898.05	4889.94	35455.79	4881.99	36130.65	4861.57	37229.33	5192.18
Students per computer	—	—	—	—	8.03	4.67	8.04	4.65	7.35	4.84	6.81	4.46	—	—	—	—
Percent of students eligible for lunch subsidy	—	—	—	—	27.66	17.90	29.20	17.99	29.42	18.12	29.07	17.44	28.35	16.89	29.01	17.25
Poverty rate	12.29	7.74	12.44	7.86	10.91	8.04	11.59	7.40	10.74	7.50	11.81	7.70	10.63	7.05	9.73	6.55
Average adjusted gross income per exemption[1] (in dollars)	13580.56	2621.25	14220.69	2790.16	14894.31	3026.38	15829.26	3194.75	16913.95	3401.20	17736.01	3500.82	18597.44	3781.75	19605.62	5359.29
Percent of grade 2 students at or above the standard on the NSRE[2] in reading	—	—	—	—	—	—	75.70	13.62	71.89	15.34	75.63	14.60	78.01	14.25	80.44	12.20
Percent of grade 4 students at or above the standard on the NSRE in mathematics concepts	—	—	17.97	16.50	—	—	32.47	19.10	37.94	17.88	38.14	18.84	42.06	19.04	45.65	20.42
Percent of grade 4 students at or above the standard on the NSRE in reading	—	—	—	—	58.17	17.89	79.19	13.22	86.12	11.00	83.04	12.70	79.30	13.26	80.29	12.08
Percent of grade 8 students at or above the standard on the NSRE in mathematics concepts	—	—	30.00	14.53	—	—	37.98	18.16	31.75	16.56	32.08	15.65	35.65	18.76	36.03	17.80
Percent of grade 8 students at or above the standard on the NSRE in reading	—	—	—	—	73.03	13.15	61.38	16.70	63.02	13.82	58.62	15.27	63.25	14.32	64.89	13.76

—Not available.

[1] From 1997–98 on, average adjusted gross income per exemption is available for all school districts (n = 248). In the remaining years, average adjusted gross income per exemption is only available for those districts that correspond directly to towns (n = 203).

[2] NSRE is the New Standards Reference Exam.

SOURCE: Vermont School Report; publications of the Vermont Department of Taxes; Vermont Department of Education files.

should be more evident when post-phase-in expenditure measures are analyzed. It is to these measures that we turn in the next section.

Results

The Distribution of Expenditures Before and After Act 60

The starting point of any evaluation of the effects of Act 60 is the choice of a measure of per pupil expenditures. When towns have been used as the unit of analysis, two measures of spending have been used in the analyses of the extent of spending inequality and the effect of Act 60 on that inequality. Heaps and Woolf (2000) used budgeted expenditures per equalized pupil. However, because many towns send or receive students for whom tuition is being paid, inequality in budgeted expenditures may overstate true spending inequality. For example, overstatement of inequality could result because budgeted expenditures per equalized pupil are based on residential pupil counts that do not include tuitioning students, resulting in artificially high per pupil numbers for districts receiving tuition students. So, as an alternative, other analysts, like Jimerson (2001) and Baker (2001), use measures of spending based on the state's calculation of local education spending per equalized pupil. Local education spending is that portion of a school district budget paid by the general state support grant, local education tax revenues, and any aid from the sharing pool when applicable. Local education spending does not include federal aid or privately donated dollars.

The starting point of any evaluation of the effects of Act 60 is the choice of a measure of per pupil expenditures.

In what follows, both measures of spending are considered since neither is a perfect indicator of the educational opportunities available to students in a town. The argument for using budgeted expenditures is that this measure includes expenditures out of not only noncategorical state aid and property tax income but also expenditures out of such diverse income sources as categorical aid for special education[15] and income from the private donations to the schools. But because of the problems created by students for whom tuition payments are being made, local education spending per pupil must also be considered.

At the school district level, the choice of expenditure measures is somewhat more clear-cut. Current expenditures per pupil measures noncapital spending; total expenditures per pupil includes current and capital spending. In the analysis that follows, both of these measures are examined. It is not possible, however, to examine the extent to which cost-adjusted spending has become more equal. Both before and after Act 60, the state aid formula recognized the fact that certain students are more costly to educate, basing aid amounts not on raw pupil counts but on equalized pupil counts. The equalized pupil count was determined by assigning weights of 1.25 to secondary students and students receiving food stamps, assigning weights of 1.2 to students with limited English proficiency, and averaging the weighted counts across 2 school years (Mathis 2001). Since these weights are ad hoc and other critical determinants of cost are not taken into account, the cost adjustments in the basic state aid formula will be imperfect (Downes and Pogue 1994). Categorical aid programs, like a small schools grant program that was established by Act 60, may help to reduce inequality in cost-adjusted aid. Nevertheless, any inequality measures presented below undoubtedly understate the extent of inequality in cost-adjusted spending, since high-cost districts are typically low-spending districts.

While the circumstances cited by the plaintiffs in *Brigham v. State* existed for many years, trends in spending inequality in the late 1980s and early 1990s undoubtedly contributed to the decision to file suit. For example, from 1989–90 to 1994–95, current expenditures per pupil had grown at an annual rate of

[15] Since categorical aid is fungible, increases in categorical aid do increase the opportunities even for those students toward whom the aid is not targeted.

3.77 percent at the top of the range. The annual growth rate at the bottom of the range was only 1.9 percent. The *Brigham* decision was handed down when the dispersion in expenditures was large and growing.

Standard inequality measures like the coefficient of variation and the Gini coefficient can both reflect the tail end of these trends and can provide an initial indication of the impact of Act 60 on spending inequality.[16] And when the town level measures of spending are used, the initial indication is that spending has become more equal post–Act 60. In particular, for budgeted expenditures in 1998–99, the coefficient of variation was 0.1317 and the Gini coefficient was 0.0728. In 2000–01, the coefficient of variation was 0.1158 and the Gini coefficient was 0.0652. These measures of inequality increased after 2000–01; in 2002–03 they equaled 0.1249 for the coefficient of variation and 0.0699 for the Gini coefficient. But both measures were still below their 1998–99 levels. And since Act 60 was already being phased-in in 1998–99, these numbers probably understate the extent to which inequality in education spending by towns has declined after the implementation of Act 60.

What equalization of educational opportunities does require is elimination of the positive correlation between expenditures and taxable wealth.

Inequality measures at the school district level tell much the same story as town level inequality measures. For example, for those school districts serving students in grades K–12, the coefficient of variation of current expenditures per pupil was 0.1280 in 1995–96. For these districts, the coefficient of variation increased to 0.1358 in 1996–97 and 0.1380 in 1997–98, the last pre–Act 60 year. In the post–Act 60 period, the coefficient of variation for current expenditures per pupil in these school districts generally declined—falling to 0.1329 in 1998–99, 0.1252 in 1999–2000, and 0.1144 in 2000–01—before increasing to 0.1339 in 2002–03.[17]

For other types of school districts, the inequality measures tend to tell the same story: fluctuating inequality pre–Act 60 and reduced inequality post–Act 60. The one exception occurs for those elementary school districts located in towns that belong to union or joint high school districts.[18] For these districts, the coefficients of variation in current expenditures per pupil were 0.4083 in 1995–96, 0.1579 in 1996–97, 0.1676 in 1997–98, 0.1834 in 1998–99, 0.1938 in 1999–2000, 0.1974 in 2000–01, and 0.1956 in 2001–02.[19] The absence of any decline in inequality in spending in the post–Act 60 period may be attributable to the ability of some of these districts to circumvent Act 60.

The stability in inequality in elementary school districts located in towns that belong to union or joint high school districts does not, by itself, indicate that goals of Act 60 have not been accomplished, since dispersion of expenditures does not imply unequal opportunities attributable to differences in taxable wealth, a reality that was recognized by the Vermont Supreme Court. For instance, dispersion in current expenditures per pupil could exist and be unrelated to property wealth if the state targeted categorical aid to districts with a greater proportion of disadvantaged students. What equalization of educational opportunities does require is elimination of the positive correlation between expenditures and taxable wealth. That this is the case is made clear in the *Brigham* decision:

[16] Expenditures were weighted by enrollment in the calculation of the inequality measures. See Murray, Evans, and Schwab (1998) for further discussion of the need for weighing by enrollment.

[17] The pattern of Gini coefficients for current expenditures per pupil in these districts is very similar. The values of the Gini coefficients were 0.0732 in 1995–96, 0.0777 in 1996–97, 0.0810 in 1997–98, 0.0778 in 1998–99, 0.0727 in 1999–2000, 0.0645 in 2000–01, and 0.769 in 2002–03.

[18] In addition to K–12 districts and elementary districts located in towns that belong to union or joint high school districts, the other large group of districts is elementary districts located in towns that tuition out their high school students.

[19] Again, the pattern of Gini coefficients for current expenditures per pupil in these districts is very similar. The values of the Gini coefficients were 0.1664 in 1995–96, 0.0889 in 1996–97, 0.0929 in 1997–98, 0.0983 in 1998–99, 0.1005 in 1999–2000, 0.1019 in 2000–01, and 0.1042 in 2001–02.

> Equal educational opportunity cannot be achieved when property-rich school districts may tax low and property-poor districts must tax high to achieve even minimum standards. Children who live in property-poor districts and children who live in property-rich districts should be afforded a substantially equal opportunity to have access to similar educational revenues. (Brigham v. State, 166 Vt. at 268)

Simple inequality measures do not tell us the extent to which Act 60 has produced a system of school financing in which the correlation between spending and wealth has been reduced. Thus, following the logic of Downes (1992), simple ordinary least squares regressions of the spending measures on measures of local resources were used to determine the extent to which Act 60 has reduced this correlation. For towns, the results of these regressions are presented in tables 2A and 2B.

Equal educational opportunity cannot be achieved when property-rich school districts may tax low and property-poor districts must tax high to achieve even minimum standards.

For the 246 towns in Vermont for which the relevant data are available, the first part of table 2A indicates that the correlation between budgeted expenditures per equalized pupil and equalized assessed valuation per pupil[20] was .515,[21] clear evidence that districts with more real property wealth did have higher per pupil expenditures prior to Act 60. Since Act 60 was already being phased-in in 1998–99, this correlation probably understates the actual strength of the relationship between expenditures and property wealth prior to the *Brigham* decision.

The remainder of the first column of table 2A shows that, while the extent of inequality in educational opportunities varies across potential measures of taxable resources, the conclusion that opportunities were unequal does not depend on the measure of taxable resources used. For example, if permanent income is taken as the measure of taxable resources and median family income is used to proxy for permanent income, the correlation between budgeted expenditures per equalized pupil and taxable resources is .295, much less than the correlation between budgeted expenditures and equalized assessed valuation but still strong.

As the discussion of table 1 indicated, after Act 60 dispersion in expenditures was reduced, even in the phase-in years. Nevertheless, dispersion remained. But the *Brigham* decision did not require equalization of expenditures; the decision required the ability to fund public education to be independent of (or negatively correlated with) taxable wealth. The second column of table 2A and both columns of table 2B provide the evidence needed to determine if Act 60 has resulted in an education financing system that satisfies the *Brigham* decision. From 1998–99 to 2002–03, the correlation between equalized assessed valuation per pupil and budgeted expenditures per equalized pupil fell from .516 to .077 and, in the latter year, was insignificant at the 10 percent level. Similar weakening in the relationship between this expenditure measure and other measures of taxable resources can be seen in table 2A. Further, in table 2B, which gives only post–Act 60 correlations between taxable resource measures and local expenditures per equalized pupil, the estimated relationship between equalized assessed valuation per pupil and local expenditures per equalized pupil is actually negative. Median family income continues to be positively related to local expenditures per equalized pupil, though this relationship does appear to be weakening over time.

In combination with the evidence on the simple distributions of expenditures, these results support the

[20] Because of data limitations, equalized assessed valuation can only be calculated for 1998–99. The 1998–99 values are used throughout this analysis. While pre–Act 60 measures of property wealth would probably be preferable, Act 60–induced changes in property values were unlikely to be apparent in 1998–99, the first year of the phase-in of Act 60.

[21] It is not possible to separate capital expenditures out of this measure of per pupil expenditures. No clear indication exists as to whether the correlation of this expenditure measure with assessed valuation overstates or understates the correlation of current expenditures with assessed valuation. Thus, some caution must be exercised in interpreting these correlations.

Table 2A. Relationships between expenditures and wealth measures for Vermont towns: 1998–2003

[Dependent variable: Budgeted expenditures per equalized pupil]

Variable	1998–99	2002–03
	Part 1	
Intercept	7099.289 (125.227)	9587.354 (103.488)
Equalized assessed valuation per pupil	0.00093 (0.00027)	0.00011 (0.00012)
R^2	0.266	0.006
Correlation coefficient	0.515	0.077
	Part 2	
Intercept	5932.714 (357.866)	8373.432 (396.410)
Median family income in 1989	0.04901 (0.01110)	0.03906 (0.01147)
R^2	0.0873	0.0413
Correlation coefficient	0.295	0.203
	Part 3	
Intercept	5786.738 (298.144)	8360.895 (398.805)
Equalized assessed valuation per pupil	0.00088 (0.00024)	0.00007 (0.00008)
Median family income in 1989	0.04086 (0.00910)	0.03821 (0.01149)
R^2	0.325	0.0439
Correlation coefficient	0.570	0.210

NOTE: Robust standard errors in parentheses.

SOURCE: Vermont School Report; publications of the Vermont Department of Taxes; Vermont Department of Education files; Heaps and Woolf (2000); Vermont Indicators Online database.

Table 2B. Relationships between expenditures and wealth measures for Vermont towns: 2000–03

[Dependent variable: Local expenditures per equalized pupil]

Variable	2000-01	2002–03
	Part 1	
Intercept	6980.417 (75.355)	7918.069 (85.572)
Equalized assessed valuation per pupil	−0.00036 (0.00011)	−0.00028 (0.00010)
R^2	0.0613	0.0403
Correlation coefficient	0.248	0.201
	Part 2	
Intercept	5419.094 (303.555)	6439.233 (400.806)
Median family income in 1989	0.04200 (0.00941)	0.03927 (0.01226)
R^2	0.0862	0.0509
Correlation coefficient	0.294	0.226
	Part 3	
Intercept	5477.803 (278.514)	6501.401 (376.359)
Equalized assessed valuation per pupil	−0.00042 (0.00015)	−0.00033 (0.00013)
Median family income in 1989	0.04658 (0.00848)	0.04345 (0.01139)
R^2	0.172	0.127
Correlation coefficient	0.415	0.356

NOTE: Robust standard errors in parentheses.

SOURCE: Vermont School Report; publications of the Vermont Department of Taxes; Vermont Department of Education files; Heaps and Woolf (2000); Vermont Indicators Online database.

view that a good faith effort has been made to satisfy the *Brigham* decision. While the correlation between taxable resources and the two expenditure measures considered here has not been reduced to zero, educational opportunities were more equal in 2002–03 than in 1998–99.[22]

When we turn to school districts as the unit of analysis, the results do not provide quite as unequivocal picture of the impact of Act 60 on the correlation between wealth measures and per pupil spending. In tables 3A and 3B, the results of regressions like those that generated the results in tables 2A and 2B are reported for the case when K–12 districts are the unit of analysis. In tables 4A and 4B, elementary school districts located in towns that belong to union or joint high school districts are the unit of analysis.[23]

When current expenditures per pupil is used as the spending measure, the correlations between spending and wealth decline for each wealth measure both for K–12 districts and for elementary school districts located in towns that belong to union or joint high school districts. If, however, total expenditures per pupil is used as the spending measure, there is not consistent evidence of a weakening in the relationship between spending and wealth. For each of the wealth measures, there is evidence of a decline in the correlation between equalized assessed value per pupil and wealth, when K–12 districts are the unit of analysis. However, for each of the three wealth measures, the correlation between total expenditures per pupil and wealth has increased for elementary school districts located in towns that belong to union or joint high school districts. These results confirm the picture created by the simple inequality measures; Act 60 has had less of an impact on inequality in educational opportunities in elementary school districts located in towns that belong to union or joint high school districts.

Student Performance Before and After Act 60

As the discussion in "School Finance Reform in Vermont," above, indicates, the *Brigham* decision focused on spending inequities. Further, the goals of Act 60 were to reduce spending inequities and to provide property tax relief. Nevertheless, the justices of the Vermont Supreme Court made clear in their decision that in their view inequities in expenditures were likely to translate into inequities in outcome:

> While we recognize that equal dollar resources do not necessarily translate equally in effect, there is no reasonable doubt that substantial funding differences significantly affect opportunities to learn. To be sure, some school districts may manage their money better than others, and circumstances extraneous to the educational system may substantially affect a child's performance. Money is clearly not the only variable affecting educational opportunity, but it is one that government can effectively equalize. (Brigham v. State, 166 Vt. at 255–56)

Money is clearly not the only variable affecting educational opportunity, but it is one that government can effectively equalize.

Thus, consideration must be given to how the distribution across districts of student performance changed after Act 60.

[22] Given the available data, it was not possible to quantify directly the strength of the correlation between expenditures and wealth measures prior to implementation of Act 60. However, the results of Baker (2001) provide an indirect indication of the strength of the correlation. In regressions that are analogous to those in part 3 of table 2B, Baker generates R^2s ranging from 0.47 to 0.51 for the school years from 1994–95 to 1998–99. Further, the highest R^2 occurs in 1998–99, the first year of the Act 60 phase-in. The implication, then, is that the correlation between expenditures and the wealth measures considered in this paper was probably strong and stable in the years leading up to Act 60.

[23] All of these regressions have been estimated in log-log form with contemporaneous measures of per pupil equalized assessed value and with adjusted gross income per exemption replacing the lagged measures used in tables 3A, 3B, 4A, and 4B. The implications of the results that are generated from these alternative specifications are the same as those reported here.

Table 3A. Relationships between expenditures and wealth measures for Vermont K–12 districts: 1997–2002

[Dependent variable: Current expenditures per equalized pupil]

Variable	1997–98	2001–02
	Part 1	
Intercept	5593.263 (232.264)	8373.759 (421.578)
Equalized assessed valuation per pupil in 1999	0.00190 (0.00059)	0.00179 (0.00133)
R^2	0.182	0.087
Correlation coefficient	0.426	0.295
	Part 2	
Intercept	4539.397 (779.732)	7606.276 (840.734)
Adjusted gross income per exemption in 1995	0.11919 (0.05448)	0.09352 (0.05480)
R^2	0.155	0.050
Correlation coefficient	0.394	0.223
	Part 3	
Intercept	5047.538 (940.288)	8219.601 (1124.209)
Equalized assessed valuation per pupil in 1999	0.00131 (0.00085)	0.00157 (0.00197)
Adjusted gross income per exemption in 1995	0.05392 (0.07972)	0.01490 (0.10701)
R^2	0.209	0.091
Correlation coefficient	0.457	0.302

NOTE: Robust standard errors in parentheses.

SOURCE: Vermont School Report; publications of the Vermont Department of Taxes; Vermont Department of Education files.

Table 3B. Relationships between expenditures and wealth measures for Vermont K–12 districts: 1997–2002

[Dependent variable: Total expenditures per equalized pupil]

Variable	1997–98	2001–02
	Part 1	
Intercept	6917.383 (483.383)	9673.285 (608.913)
Equalized assessed valuation per pupil in 1999	0.00122 (0.00087)	0.00128 (0.00169)
R^2	0.021	0.021
Correlation coefficient	0.145	0.145
	Part 2	
Intercept	5464.105 (1577.796)	8967.333 (906.447)
Adjusted gross income per exemption in 1995	0.12711 (0.10495)	0.07108 (0.05652)
R^2	0.048	0.012
Correlation coefficient	0.219	0.112
	Part 3	
Intercept	5448.176 (1867.612)	9457.480 (1263.456)
Equalized assessed valuation per pupil in 1999	–0.00004 (0.00132)	0.00125 (0.00270)
Adjusted gross income per exemption in 1995	0.12915 (0.14738)	0.08242 (0.13496)
R^2	0.048	0.024
Correlation coefficient	0.218	0.155

NOTE: Robust standard errors in parentheses.

SOURCE: Vermont School Report; publications of the Vermont Department of Taxes; Vermont Department of Education files.

Table 4A. Relationships between expenditures and wealth measures for Vermont elementary districts located in towns that do not tuition out high school students: 1997–2002

[Dependent variable: Current expenditures per equalized pupil]

Variable	1997–98	2001–02
	Part 1	
Intercept	5512.449 (170.779)	7986.672 (237.922)
Equalized assessed valuation per pupil in 1999	0.00092 (0.00023)	0.00134 (0.00028)
R^2	0.170	0.157
Correlation coefficient	0.412	0.396
	Part 2	
Intercept	5116.117 (505.333)	8018.435 (829.521)
Adjusted gross income per exemption in 1995	0.07212 (0.03679)	0.06213 (0.05965)
R^2	0.032	0.010
Correlation coefficient	0.179	0.099
	Part 3	
Intercept	5199.331 (481.227)	8152.715 (756.184)
Equalized assessed valuation per pupil in 1999	0.00083 (0.00023)	0.00134 (0.00030)
Adjusted gross income per exemption in 1995	0.02712 (0.03741)	−0.01048 (0.05583)
R^2	0.162	0.151
Correlation coefficient	0.402	0.388

NOTE: Robust standard errors in parentheses.

SOURCE: Vermont School Report; publications of the Vermont Department of Taxes; Vermont Department of Education files.

Table 4B. Relationships between expenditures and wealth measures for Vermont elementary districts located in towns that do not tuition out high school students: 1997–2002

[Dependent variable: Total expenditures per equalized pupil]

Variable	1997–98	2001–02
	Part 1	
Intercept	6467.415 (230.387)	8859.059 (273.432)
Equalized assessed valuation per pupil in 1999	0.00091 (0.00022)	0.00147 (0.00028)
R^2	0.065	0.118
Correlation coefficient	0.254	0.344
	Part 2	
Intercept	5635.405 (764.984)	8026.746 (1025.426)
Adjusted gross income per exemption in 1995	0.10323 (0.05418)	0.13226 (0.07569)
R^2	0.024	0.028
Correlation coefficient	0.155	0.169
	Part 3	
Intercept	5717.045 (737.757)	8164.198 (946.995)
Equalized assessed valuation per pupil in 1999	0.00081 (0.00024)	0.00137 (0.00031)
Adjusted gross income per exemption in 1995	0.05908 (0.05555)	0.05793 (0.07331)
R^2	0.071	0.122
Correlation coefficient	0.266	0.349

NOTE: Robust standard errors in parentheses.

SOURCE: Vermont School Report; publications of the Vermont Department of Taxes; Vermont Department of Education files.

A crude indication of the impact of Act 60 on student performance is given by table 5, which presents correlations in 1995–96 and 2001–02 between some of the district characteristics summarized in table 1. The correlations between student performance and all available measures of the resources allocated towards education have weakened in the post–Act 60 period. The starkest example of the weakening of these relationships is the decline from 1995–96 to 2001–02 of the relationship between current expenditures per pupil and the percent of eighth-graders at or above the standard for the concepts portion of the New Standards Reference Exam (NSRE) in mathematics.[24] A more systematic assessment of the impact of Act 60 can be based on the results in table 6, which gives a few typical event-analysis regressions that are similar in flavor to those in Downes and Figlio (2000).[25]

Because they include controls for district-specific effects and because they are based on a functional form that explicitly accounts for the reality that the share of students meeting the standard must range between 0 and 1, regressions like those in table 6 provide the most convincing estimates of the impact of Act 60. Further, because in these regressions the impact of Act 60 is allowed to vary with pre–Act 60 spending levels or pre–Act 60 property wealth, the regressions provide a direct indication of the extent to which the link between wealth and performance has changed post–Act 60. And, what is apparent from these regressions, and from a number of regressions in which other outcome measures are used as the dependent variable, is that there is some evidence that the gaps in performance between high-spending and low-spending districts and between high-wealth and low-wealth districts have, *ceteris paribus*, declined post–Act 60. In these regressions, the coefficient on the interaction between the Act 60 dummy and the pre–Act 60 spending or pre–Act 60 wealth is never positive and significant. And, as can be seen in table 6, these coefficients are frequently negative and significant.

Care must be taken, however, not to make too much of the declines in inequality. The coefficients on the interactions are not consistently negative and significant. Further, when these coefficients are significant, they are quantitatively small. For example, the coefficient in the first column of table 6 implies that, *ceteris paribus*, the difference between the shares of fourth-graders at or above the standard for a school district with spending one standard deviation below the mean in 1994–95 and a school district with spending one standard deviation above the mean in 1994–95 would decline by 0.0021 if each district had the mean number of test takers in 2000–01. It seems unlikely that such small declines in dispersion in performance justify a major policy intervention like Act 60.

Act 60 has dramatically reduced dispersion in education spending by weakening the link between spending and property wealth.

Concluding Remarks

Act 60 represents a dramatic change in the system of education financing in a state with a history of a demographically stable student population. As a result, Act 60 may well provide an unparalleled opportunity to assess the impact of a significant finance reform on a state's education system. This paper represents a first cut at just such an assessment.

All of the evidence cited in this paper supports the conclusion that Act 60 has dramatically reduced dispersion in education spending and has done this by weakening the link between spending and property wealth. Further, the regressions presented in this paper offer some evidence that student performance has become more equal in the post–Act 60 period. And no results support the conclusion that Act 60 has contributed to increased dispersion in performance.

[24] Jimerson (2001) observes a similar decline in the correlation between equalized assessed value per pupil and the percent of fourth-graders at or above the standard for the NSRE.

[25] A traditional event-analysis approach is preferable to the production function approach used by Flanagan and Murray (2002) because the production function approach can only provide accurate estimates of the impact of the policy changes if all of the effects of the changes work through changes in measured inputs and if all of the changes in inputs are attributable to the policy changes.

Table 5. Correlations between selected characteristics for Vermont school districts, 1995–96 and 2001–02

Variable	Current expenditures per pupil	Students per classroom teacher	Average teacher salary	Poverty rate	Average adjusted gross income per exemption (from tax returns)	Percent of adults with college degree (1990)	Percent of grade 4 students at or above the standard on the NSRE[1] in mathematics concepts	Percent of grade 8 students at or above the standard on the NSRE[1] in mathematics concepts
				1995–96				
Current expenditures per pupil	1.0000							
Students per classroom teacher	–0.0511	1.0000						
Average teacher salary	0.2656	–0.2085	1.0000					
Poverty rate	–0.1021	–0.0102	–0.1504	1.0000				
Average adjusted gross income per exemption (from tax returns)	0.2286	–0.0522	0.5196	–0.5655	1.0000			
Percent of adults with college degree (1990)	0.1682	–0.0081	0.3916	–0.5683	0.7964	1.0000		
Percent of grade 4 students at or above the standard on the NSRE[1] in mathematics concepts	0.0827	0.0244	0.0731	–0.1993	0.1966	0.2135	1.0000	
Percent of grade 8 students at or above the standard on the NSRE[1] in mathematics concepts	0.2115	–0.0461	0.2283	–0.3585	0.3195	0.3812	0.1876	1.0000
				2001–02				
Current expenditures per pupil	1.0000							
Students per classroom teacher	–0.3108	1.0000						
Average teacher salary	0.0508	0.3572	1.0000					
Poverty rate	–0.0541	–0.0955	–0.2405	1.0000				
Average adjusted gross income per exemption (from tax returns)	0.1182	0.2716	0.5416	–0.5326	1.0000			
Percent of adults with college degree (1990)	0.2359	0.1540	0.4235	–0.5772	0.7901	1.0000		
Percent of grade 4 students at or above the standard on the NSRE[1] in mathematics concepts	0.0931	0.1022	0.1623	–0.3493	0.3077	0.3884	1.0000	
Percent of grade 8 students at or above the standard on the NSRE[1] in mathematics concepts	0.1885	0.1614	0.3146	–0.3562	0.4936	0.5084	0.3378	1.0000

[1] NSRE is the New Standards Reference Exam.

SOURCE: Vermont School Report; publications of the Vermont Department of Taxes; Vermont Department of Education files.

Table 6. Generalized Linear Model estimates of impact of Act 60 on student performance—fixed effects estimates:[1] 1995–2002

[Dependent variable: Number of test-takers at or above the standard in mathematics concepts on the NSRE]

Variable	Fourth-Graders		Eighth-Graders	
	Specification 1	Specification 2	Specification 1	Specification 2
Dummy variable indicating post–Act 60	−12.6896 (0.9788)	0.3965 (0.3103)	−12.3158 (1.1338)	−1.7567 (0.3484)
Interaction of post–Act 60 dummy with per pupil equalized assessed valuation—1998	—	−0.00000001 (0.00000012)	—	0.00000013 (0.00000008)
Interaction of post–Act 60 dummy and current expenditures per pupil—1995	−0.00003 (0.00001)	—	−0.00002 (0.00002)	—
Poverty rate	0.0046 (0.0143)	0.0059 (0.0145)	0.0029 (0.0163)	0.0065 (0.0199)
Adjusted gross income per exemption	−0.00001 (0.00001)	−0.00001 (0.00001)	0.00003 (0.00002)	0.00003 (0.00002)
Dummy variable indicating 1995–96 school year	−14.0105 (1.2344)	−0.7150 (0.2912)	−12.3880 (1.1230)	−1.7779 (0.3189)
Dummy variable indicating 1997–98 school year	−13.1792 (1.0415)	0.0874 (0.3076)	−12.1258 (1.1304)	−1.5372 (0.3351)
Dummy variable indicating 1999–2000 school year	−0.0108 (0.0463)	−0.0182 (0.0463)	−0.0078 (0.0547)	−0.0018 (0.0543)
Dummy variable indicating 2000–01 school year	0.1823 (0.0563)	0.1827 (0.0554)	0.1672 (0.0755)	0.1719 (0.0748)
Dummy variable indicating 2001–02 school year	0.3295 (0.0840)	0.3285 (0.0830)	0.2321 (0.1002)	0.2364 (0.0987)
Log of likelihood function	−3109.6629	−3159.0275	−1923.2455	−1955.8046
Number of observations	1182	1190	690	701

—Not available.

[1] The constant is omitted from each specification. The omitted school year is 1998–99.

NOTE: Asymptotic standard errors robust to heteroskedasticity and within group correlation in parentheses. NSRE is the New Standards Reference Exam.

SOURCE: Vermont School Report; publications of the Vermont Department of Taxes; Vermont Department of Education files.

By themselves, these results may provide useful information for policymakers contemplating Act 60–style reforms. But the value of these results may well increase dramatically when taken together with the results of Duncombe and Johnston (2002) and of Flanagan and Murray (2002). What is striking is the similarity across studies in the estimated achievement effects. Pre-finance reform data on student test scores are not available to Duncombe and Johnston; they find no evidence that a diminishment in the dispersion in performance is apparent when examining post-finance-reform test scores. They also document some recent relative improvement in dropout rates in high poverty districts, though they also find increased dispersion in dropout rates when comparing pre- and post-finance-reform data.

The bottom line of Duncombe and Johnston's analysis of dropout rates is that reform has resulted in small relative improvements. Flanagan and Murray reach conclusions similar to those reached in this paper—post-reform dispersion in schooling outcomes has declined, but this decline in dispersion has been small. The results presented above indicate that, in Vermont, there have been, at most, small relative improvements in the test performance of fourth- and eighth-graders in those school districts with lower pre-reform per pupil spending and per pupil property wealth. Flanagan and Murray find that relative increases in post-reform spending were translated into relative gains in post-reform test performance, but these gains were quantitatively small. Somewhat surprisingly, then, the results of these new case studies tend to echo the results of the earlier work on California. Thus far, the case studies have confirmed a conclusion that was reached by many of the researchers who executed national-level analyses: the types of finance reforms that have been implemented in response to court orders appear to have little, if any, impact on the distribution of student test performance.

References

Baker, B.D. (2001). Balancing Equity for Students and Taxpayers: Evaluating School Finance Reform in Vermont. *Journal of Education Finance, 26*(Spring): 239–248.

Bènabou, R. (1996). Equity and Efficiency in Human Capital Investment: The Local Connection. *Review of Economic Studies, 63*(April): 237–264.

Brigham v. State, 166 Vt. 246, available at http://dol.state.vt.us/gopher_root1/supct/166/96-502.op

Burkett, E. (1998, April 26). Don't Tread on My Tax Rate. *The New York Times Magazine,* 42–45.

Card, D., and Payne, A.A. (2002). School Finance Reform, the Distribution of School Spending, and the Distribution of Student Test Scores. *Journal of Public Economics, 83*(January): 49–82.

Courant, P.N., and Loeb, S. (1997). Centralization of School Finance in Michigan. *Journal of Policy Analysis and Management, 16*(Winter): 114–136.

Downes, T.A. (1992). Evaluating the Impact of School Finance Reform on the Provision of Public Education: The California Case. *National Tax Journal, 45*(December): 405–419.

Downes, T.A., and Figlio, D.N. (1999). *What Are the Effects of School Finance Reforms? Estimates of the Impact of Equalization on Students and on Affected Communities.* Unpublished manuscript, Tufts University, Medford, MA.

Downes, T.A., and Figlio, D.N. (2000). *School Finance Reforms, Tax Limits, and Student Performance: Do Reforms Level-Up or Dumb Down?* Unpublished manuscript, Tufts University, Medford, MA.

Downes, T.A., and Pogue, T.F. (1994). Accounting for Fiscal Capacity and Need in the Design of School Aid Formulas. In J.E. Anderson (Ed.), *Fiscal Equalization for State and Local Government Finance.* New York: Praeger.

Downes, T.A., and Shah, M. (1995, June). *The Effect of School Finance Reform on the Level and Growth of Per Pupil Expenditures* (Tufts University Working Paper No. 95-94). Medford, MA: Tufts University.

Duncombe, W., and Johnston, J.M. (2002). *Is Something Better Than Nothing? An Assessment of School Finance Reform in Kansas* (Working Paper). Syracuse, NY: Center for Policy Research, Maxwell School of Citizenship and Public Affairs, Syracuse University.

Evans, W.N., Murray, S., and Schwab, R.M. (1997). Schoolhouses, Courthouses, and Statehouses after *Serrano*. *Journal of Policy Analysis and Management 16*(Winter): 10–31.

Evans, W.N., Murray, S., and Schwab, R.M. (1999). The Impact of Court-Mandated School Finance Reform. In H.F. Ladd, R. Chalk, and J.S. Hansen (Eds.), *Equity and Adequacy in Education Finance: Issues and Perspectives.* Washington, DC: The National Academies Press.

Fernandez, R., and Rogerson, R. (1997). Education Finance Reform: A Dynamic Perspective. *Journal of Policy Analysis and Management 16*(Winter): 67–84.

Fernandez, R., and Rogerson, R. (1998). Public Education and Income Distribution: A Dynamic Qualitative Evaluation of Education Finance Reform. *American Economic Review 88*(September): 813–833.

Flanagan, A., and Murray, S. (2002). *A Decade of Reform: The Impact of School Reform in Kentucky.* Unpublished manuscript, RAND, Washington, DC.

Heaps, R., and Woolf, A. (2000, September). *Evaluation of Act 60: Vermont's Education Financing Law.* Unpublished manuscript, Northern Economic Consulting, Inc., Burlington, VT.

Hoxby, C.M. (2001). All School Finance Equalizations Are Not Created Equal: Marginal Tax Rates Matter. *The Quarterly Journal of Economics, 116*(November): 1189–1231.

Husted, T.A., and Kenny, L.W. (2000). Evidence on the Impact of State Government on Primary and Secondary Education and the Equity-Efficiency Trade-Off. *Journal of Law and Economics, 43*(1): 285–308.

Jimerson, L. (2001, February). A Reasonably Equal Share: *Educational Equity in Vermont: A Status Report—Year 2000–2001* (Report of the Rural School and Community Trust Policy Program). Washington, DC: Rural School and Community Trust.

Manwaring, R.L., and Sheffrin, S.M. (1997). Litigation, School Finance Reform, and Aggregate Educational Spending. *International Tax and Public Finance, 4*(2): 107–27.

Mathis, W.J. (1995). Vermont. In S.D. Gold, D.M. Smith, and S.B. Lawton (Eds.), *Public School Finance Programs of the United States and Canada, 1993–94.* Albany, NY: The Nelson Rockefeller Institute of Government, State University of New York.

Mathis, W.J. (1998, July). *Act 60 and Proposition 13.* Montpelier, VT: Concerned Vermonters for Equal Educational Opportunity. Available at http://www.act60works.org/oped1.html

Mathis, W.J. (2001). Vermont. In C.C. Sielke, J. Dayton, C.T. Holmes, and A.L. Jefferson (Comps.), *Public School Finance Programs of the U.S. and Canada: 1998–99* (NCES 2001–309). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

McClaughry, J. (1997, December). *Educational Financing Lessons From California* (Ethan Allen Institute Commentary). Concord, VT: Ethan Allen Institute. Available at http://www.ethanallen.org/commentaries/1997/educatingfinancial.html

McClaughry, J. (2001, January). *Schoolchildren First: Replacing Act 60* (Ethan Allen Institute Policy Brief 007). Concord, VT: Ethan Allen Institute. Available at http://www.ethanallen.org.html

Murray, S.E., Evans, W.N., and Schwab, R.M. (1998). Education Finance Reform and the Distribution of Education Resources. *American Economic Review, 88*(4):789–812.

Nechyba, T.J. (1996, June). *Public School Finance in a General Equilibrium Tiebout World: Equalization Programs, Peer Effects and Private School Vouchers* (NBER Working Paper No. w5642). Cambridge, MA: National Bureau of Economic Research.

Nechyba, T.J. (2000). Mobility, Targeting and Private School Vouchers. *American Economic Review, 90*(1): 130–46.

Silva, F., and Sonstelie, J. (1995). Did *Serrano* Cause a Decline in School Spending? *National Tax Journal, 48*(2): 199–215.

Shopping for Evidence Against School Accountability

Margaret E. Raymond
Eric A. Hanushek
Stanford University

About the Authors

Margaret E. Raymond is director of CREDO at the Hoover Institution of Stanford University. CREDO provides impartial evaluation of educational programs and policies. She has published independent evaluations of Teach for America, charter schools, and California's accountability system. She can be contacted at macke@stanford.edu.

Eric A. Hanushek is the Paul and Jean Hanna Senior Fellow at the Hoover Institution of Stanford University, chair of the executive committee for the Texas Schools Project, and a research fellow of the National Bureau of Economic Research. He has written extensively about the economics and finance of schools. He can be contacted at hanushek@stanford.edu.

Shopping for Evidence Against School Accountability

Margaret E. Raymond
Eric A. Hanushek
Stanford University

Accountability has been a central feature of educational policy in a number of states since the 1990s. In part because of the perceived success of accountability in the states where it was initially tried, federal law introduced mandatory reporting and accountability through the No Child Left Behind Act of 2001. Yet not everybody is happy with school accountability. Its opponents continue to aggressively search for evidence that testing and accountability do not work—or, better, that they are actually harmful. The hope of the anti-accountability forces is that they can stop testing before it is fully in place and before rollbacks would be impossible.

The window of opportunity to cripple or stop testing is narrowing over time, so it is not surprising that hasty reports based on biased research should appear. Nor is it surprising that these reports are given attention by parties who are unschooled in the requirements of good research. Perhaps we could disregard these events if the policies themselves were unimportant or if public exposure to poor quality studies had no effect on the ultimate decisions about them. But that is not the case. Since testing and accountability represent the cornerstone of current school reform efforts, it is essential that we apply rigorous standards of evidence and of scientific method to the analysis of accountability policy. The impact of testing and accountability is perhaps the most important issue facing school policymakers today. Even though accountability, by itself, does not say anything about how to organize an effective school, measures of school performance provide a standardized construction of information needed to forge through the bewildering array of "answers" to the question of how to improve our schools. While it is certainly reasonable to question the effectiveness of particular accountability systems and the policy of accountability in general, little thought has been given to the scientific standards of evidence that ought to apply to research and evaluation aimed at informing or influencing the policy process in this important area.

Assessing the impact of state accountability is clearly difficult. Policies have been in place for a limited amount of time. All states but one have adopted a system in one form or another. Not all accountability systems are the same. When put in place, they apply to all schools within entire states, limiting relevant variation to differences across states. This means that we have lost forever the chance to test whether accountability systems are superior to what states had before. Finally, accountability systems are just one of many ways in which states tend to differ. These factors do not imply that gathering evidence about the effects of accountability is impossible. They simply reinforce the need to apply strict scientific methods to ensure that uncertainty is reduced as far as possible.

Bad news about accountability gets an undue amount of media coverage. First, the anti-accountability forces trumpet any possible scrap of data that might be portrayed as generalizable evidence against routine testing and accountability. Second, researchers reinforce this by their popular search for unintended consequences of government actions. Finally, the press, looking for both controversy and balance in reporting, tends to cite any study—no matter what its scientific quality—to show the evenhandedness of its reporting.

What do we know to date? The existing evidence on state accountability systems indicates that their use leads to improvement of student achievement. States that introduced accountability systems during the 1990s tended to show more rapid achievement gains when compared to states that did not introduce such measures. Along with general improvement, there also appear to be instances of unintended consequences—such as increased special education placement or outright cheating—at the time of introduction, but there is no evidence that this continues over time. Looking across states, we also know that attaching stakes to performance on tests yields better performance. Though still preliminary, these findings rest on rigorous analytic techniques, providing policymakers the most reliable evidence yet available.

> ***The existing evidence on state accountability systems indicates that their use leads to improvement of student achievement.***

What do we not know to date? Plenty. We do not know which general designs of accountability systems work best, or even the best underlying content standards for achievement. Nor do we know the optimal way to attach rewards and punishments to performance. Who should be judged by what scores? These are things that will take time to discover, but there is no way to get from here to there without a systematic approach to future policy enhancements and continued rigorous evaluation of their effects.

Evidence About Existing Accountability Systems

Over the past decade, states have devised diverse accountability systems that differ by choice of test, grades monitored, subjects tested, and performance requirements. Direct comparison of state against state based on state accountability system information is therefore problematic; a common but independent standard of comparison is needed. One source of information on performance, however, offers some possibility for analysis. The National Assessment of Educational Progress (NAEP), the "Nation's Report Card," provided performance information for states during the 1990s. While not designed as a national test, these examinations provide a highly respected and consistent tracking of student performance across grades and time. Since scores are not reported for individuals or schools, there is no incentive to prep for them or to cheat on them. We have used these performance measures to assess the impacts of state accountability systems.

Education is the responsibility of state governments, and states have gone in a variety of directions in the regulation, funding, and operation of their schools. As a result, it is difficult to assess the impacts of individual policies without dealing with the potential impacts of coincidental policy differences.[1]

The basic analysis focuses on growth of student achievement across grades.[2] If the impacts of stable state policies enhance or detract from the educational process in a consistent manner across grades, concentrating on achievement growth implicitly allows for stable state policy influences and permits analysis of the introduction of new state accountability policies.

[1] Hanushek, Rivkin, and Taylor (1996) discuss the relationship between model specification and the use of aggregate state data. The development here builds on the prior estimation in Hanushek and Somers (2001) and the details of the model specification and estimation can be found there.

[2] Here we summarize the results of the analysis in Hanushek and Raymond (2003a, 2003b).

The NAEP testing measured math performance of fourth-graders in 1992 and 1996 and of eighth-graders 4 years after each of these assessments. While the students are not matched, following the same cohort acts to eliminate a variety of potentially confounding achievement influences. We also supplement the raw NAEP data by considering differences in parental education levels and in school spending across these states. Our analysis of achievement relies on growth in achievement in reading and math between fourth- and eighth-graders over the relevant 4-year period, e.g., growth in achievement from fourth grade in 1996 to eighth grade in 2000. Our sample is all states for which the relevant NAEP scores are available.

The potential effects of accountability systems clearly depend on when and where these systems were introduced. Table 1 describes the time path of introduction of accountability systems across states by reference to the length of time that accountability systems have been operating in different states. For these purposes, we define accountability systems as those that relate student test information to schools and either simply report scores or provide rewards and sanctions.[3] By looking at accountability systems in 1996, it is clear that much of the movement to accountability is very recent. In 1996, just 10 states had already introduced active accountability systems, while by 2000 only 13 states had yet to introduce active systems.[4]

We rely on statistical analyses of differences in NAEP growth across states to infer the impact of introducing state accountability. Because a differing set of about 40 states participated in the NAEP testing in each of the years, the amount of evidence is limited. Nonetheless, state accountability systems uniformly have a significant impact on growth in NAEP scores, while other potential influences—spending and parental education levels—do not.

Figure 1 summarizes the impact of existing state systems by tracking the gains in mathematics between 1996 and 2000 for the typical student who progresses from fourth to eighth grade under different systems. These expected gains, calculated from regression analyses of scores on NAEP, illustrate the impact of testing and reporting across states.[5] States were classified according to the type of accountability system they had in place at the time of the NAEP test. (A state's classification could change between the two test years if its accountability system had been newly adopted or changed in the interim.) The typical student in a state without an accountability system of any form would see a 0.7 percent increase in the proficiency scores between fourth and eighth grades. States with "report card" systems display test performance and other factors but do not attach sanctions and rewards to the information. In many ways, these systems simply serve a public disclosure function. Just this reporting moves

[3] We do not include states that place rewards or sanctions ("high-stakes") just on students, for example through use only of a required graduation exam. The school accountability systems are most relevant for No Child Left Behind, but this restriction introduces some differences between our analysis and the analysis of Amrein and Berliner (2002) that is analyzed below.

[4] In all analyses, the universe includes 50 states plus the District of Columbia. Nonetheless, not all states participate in the NAEP exams each year, and the samples fall to around 35 in each year.

[5] The details of these estimates can be found in Hanushek and Raymond (2003a). The results pool data on NAEP mathematics gains over both the 1992–96 and 1996–2000 periods.

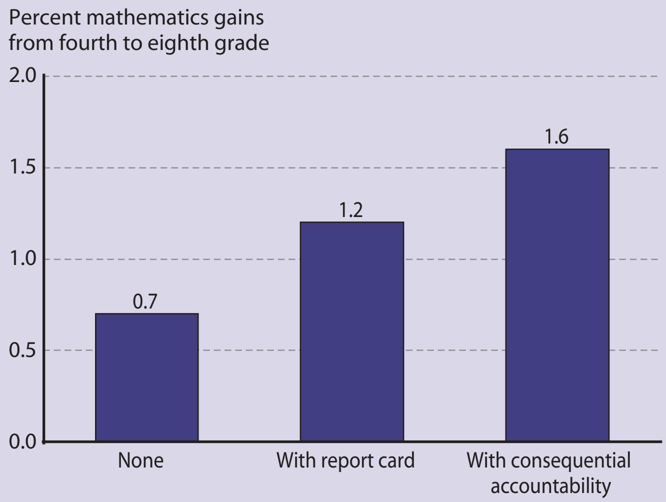
Table 1. Distribution of states with consequential accountability or reporting system: 1996 and 2000

	Number of states	
	1996	2000
No system	41	13
System in place	10	38

NOTE: Distribution includes Washington, DC.

SOURCE: Fletcher and Raymond (2002).

Figure 1. Estimated effects of state accountability systems on gains between fourth grade and eighth grade for National Assessment of Educational Progress (NAEP) mathematics scores: 1996–2000



SOURCE: Author calculations from Hanushek and Raymond (2003a, 2003b).

the expected gain to 1.2 percent. Finally, states that provide explicit scores for schools and that attach sanctions and rewards (what we call "consequential accountability" systems) obtained a 1.6 percent increase in mathematics proficiency scores. In short, testing and accountability as practiced have led to significant gains in student performance over that expected without formal systems.

A complementary analysis by Carnoy and Loeb (2002), while not considering the timing of the introduction of accountability, includes a rating of the stringency of the accountability system that is finer grained than the two categories we employ. It also adds information about student stakes and accountability. Carnoy and Loeb's findings reinforce the present analysis that accountability increases NAEP performance. A variety of other systematic studies of accountability systems within states and local school districts have also investigated what happens when accountability systems are introduced. While we describe the evidence in detail elsewhere (Hanushek and Raymond 2003a, 2003b), it generally supports two conclusions. First, improvements in available measures of student performance occur after the introduction of an accountability system. Second, other short-run changes—such as increases in test exclusions or explicit cheating—are observed. In other words, some unintended consequences often tend to accompany the introduction of accountability, although as of now there is little evidence suggesting that these influences continue over time.

We ourselves have looked explicitly at state differences in special education placement rates and whether they are related to accountability systems. For the period 1995–2000, a time of large change in the use of accountability systems, we see no evidence that increased special education placement is a reaction to accountability systems (Hanushek and Raymond 2003a, 2003b). This analysis does, however, show why some could mistakenly conclude that accountability has an impact: overall special education placement increases within states over this time period, so the introduction of accountability systems in the middle of the period can look like it influences placement.

Carnoy and Loeb (2002) also investigate the impacts of accountability on grade retention and graduation. They demonstrate that there is no discernible negative effect on retention and graduation.

The set of scientific studies of accountability has been presented at a range of scientific conferences, and many have undergone peer review for journal publication. In fact, because of the importance of the topic, the Kennedy School at Harvard held an entire conference on accountability in June 2002, and Brookings published the papers in 2003 (Peterson and West 2003).

The Allure of Counter-Evidence

In late 2002, Amrein and Berliner, hereafter AB, produced a study on the impact of high-stakes accountability systems that garnered considerable attention (AB 2002).[6] Their analysis of 28 states considers the effects on state-specific NAEP scores and college entrance examination measures in the period following adoption of a high-stakes accountability program.[7] Their analysis concludes "there is inadequate evidence to support the proposition that high-stakes tests . . . increase student achievement" (AB 2002, p. 57). The press release that describes the report goes further: "The Berliner-Amrein analyses suggest that, as indicated by student performance on independent measures of achievement, high-stakes tests may inhibit the academic achievement of students, not foster their academic growth."

Because of the importance of the topic, the Kennedy School at Harvard held an entire conference on accountability in June 2002.

A closer look at the research, however, shows it to be fatally flawed both in design and in execution, rendering the conclusions irrelevant. We consider only the effects of accountability systems on NAEP scores in the 26 states that AB record as having adopted grade school high-stakes tests.[8]

It is difficult to ascertain from the main text or the technical appendixes exactly what procedures and definitions AB employed. AB's methodology seems best described as a "pseudo-trend analysis" with, at times, absent baseline data.[9] Given the fact that state-level NAEP data on the math and reading tests are available only for at most four data points, AB essentially were confined to performing case studies of individual states.[10] They purport to examine the change in scores before and after the accountability system was adopted in each state—thus using each state as a control for itself. To give some independent context for these differences, it appears they also generally compared the state change to the change that was observed for the nation as a whole. States were coded as increasing on a particular test if the gains in average test results exceeded the national average change, or coded as decreasing in the opposite case. Finally, all scores were then considered in relation to the relative

---

[6] This study is described as having been completed for the Great Lakes Center for Education Research and Practice, a Michigan-based think tank. That organization, which is solely financed by National Education Association State Education Affiliate Associations from Illinois, Indiana, Michigan, Minnesota, Ohio, and Wisconsin, in turn describes a key element of its mission as being to "connect with like-minded organizations to partner on key education initiatives."

[7] We have not assessed their identification and timing of high-stakes testing, which apparently can relate both to school stakes and individual student stakes.

[8] Georgia and Minnesota only adopted high school exit requirements, the subject of AB's technical appendix. There are also strong reasons to question their analysis of high school level performance, given the looser degree of correspondence between high school exit requirements and college entrance test results, but that discussion necessarily gets into other issues and only distracts from the key linkages to state accountability that we emphasize here.

[9] For example, they most frequently say in the write-ups for individual states things like "After stakes were attached to tests in Maryland, grade 4 math achievement decreased" (p. 28). But, since fourth-grade NAEP scores in Maryland, like those in *all* of their high-stakes tests except Delaware in 1992–96, increased in every test year, we infer that they really meant to describe a comparison with the average national changes.

[10] Note that reading and math were tested in different years during the 1990s and that many states did not participate in all four waves of NAEP testing.

change in exclusion rates between the national average and the individual states. Where states' exclusion rates exceeded the national average, AB hypothesize that scores should rise because of these exclusions. Thus, whenever exclusion rates moved in the same direction as the observed NAEP test results, they considered the score change contaminated (regardless of the magnitudes involved) and eliminated the state from further consideration as "Unclear."[11] Finally, among states that remained (between 8 and 12 depending on the particular NAEP test), they examined the proportion of states with increases versus those with decreases relative to the national average. Based on this approach, they concluded that "67 percent of the states posted overall decreases in NAEP math grade 4 . . . 63 percent of the states posted increases in NAEP math grade 8 . . . and 50 percent of the states posted increases in NAEP reading grade 4 as compared to the nation after high-stakes tests were implemented." (AB 2002, p. 56)

AB violate the first principle of social science research—the need to control for the condition of interest. They used the 26 states with high-stakes accountability systems and limited their analysis to those states alone. The natural comparison group, however, is the states that had not adopted accountability systems. Such a comparison, which offers some insights into the impact of high-stakes testing as opposed simply to variations among states with high-stakes systems, yields starkly different results than their suggested interpretations.[12] In fact, their results are completely reversed, putting the evidence in line with that previously discussed.

Table 2 simply compares fourth- and eighth-grade NAEP test score gains for the states AB identify as implementing high-stakes testing with those that were not so identified.[13] For either the entire 1992–2000 period or the later period of 1996–2000, the average gain in math for high-stakes states significantly exceeds that for the remaining tested states. The difference in performance is always statistically significant at conventional levels (a nuance that AB never even mention in their 236 pages of analysis).[14]

AB highlight changes in exclusion rates from test taking as a possible influence on state test scores, and differences in exclusions between high-stakes states and others could influence the performance differentials shown. Indeed, many people have suggested that a

[11] In reality, they do not even appear consistent on this, and they violate their own coding scheme more than once. Take, for example, West Virginia, where they state: "Overall NAEP math grade 4 scores increased at the same time the percentage of students exempted from the NAEP increased. Overall, after stakes were attached to tests in West Virginia, grade 4 math achievement **decreased**." [their emphasis]

[12] While we reproduce their analysis with a larger set of observations, this should not be construed as an endorsement of the analytical approach. More rigorous tools yield more reliable results. We follow their lead in order to show how their answers would have differed had they applied their own approach correctly.

[13] Note that for each of the comparisons data are available for 34 to 36 states with between 18 and 20 being in the AB high-stakes sample. The limited number reflects the varying participation of states in the NAEP testing.

[14] Statistical testing is done to guard against changes in test performance that simply reflect random score differences that do not represent true differences in student performance. Such random differences could, for example, reflect chance differences in the tested population, small changes in question wording, or events specific to the testing in a given year and given state. In their subsequent defense of their analysis, AB assert that such testing is unnecessary and may even be inappropriate, but this assertion is obviously incorrect (AB 2003).

Table 2. Average gains in National Assessment of Educational Progress (NAEP) mathematics scores, by Amrein-Berliner (AB) high-stakes states versus other states: 1992–2000

	Change in fourth-grade NAEP mathematics scores		Change in eighth-grade NAEP mathematics scores	
	1992–2000	1996–2000	1992–2000	1996–2000
AB high-stakes states	9.2	4.2	8.8	4.5
Other states	3.8	2.3	4.0	1.7
High-stakes advantage	5.3	1.9	4.8	2.8
Statistical significance	$p<.001$	$p<.04$	$p<.003$	$p<.02$

SOURCE: Author calculations.

consequence of the introduction of high-stakes testing is an increase in test exclusions.

The hypothesized effect of accountability on test exclusions does not appear important in explaining the aggregate accountability results. For the nation as a whole, exclusion rates on the eighth-grade NAEP math tests were the same in 2000 as in 1992, while the fourth-grade exclusions over that time period fell slightly. Table 3 shows evidence for the NAEP exclusion rates for the 1992–2000 period for the high-stakes and non-high-stakes states. While the change in exclusion rates over the 1990s is slightly higher for high-stakes states in the testing of eighth-grade mathematics, it is slightly lower for fourth-grade mathematics when compared to other states. But neither difference in average exclusion by accountability status is statistically significant.

We also standardize the achievement gains for observed changes in exclusions through regression analysis. Interestingly, while changes in exclusion rates are significantly related to changes in eighth-grade scores, they are not significantly related to changes in fourth-grade scores—underscoring the need to analyze central maintained hypotheses. Table 4 compares such adjusted estimates of the achievement gain advantage of high-stakes tests to the previously unadjusted differences. Again, there are small effects on the estimated impact of high-stakes testing on gains, but in all cases states that introduce high-stakes testing outperform those that do not by a statistically significant margin. In sum, the previous estimates are not driven by test exclusions.

AB's choice of the pseudo-trend design is even more mysterious when one considers that it could not be applied squarely to their sample. In eight states—Colorado, Indiana, Louisiana, New Jersey, New Mexico, Oklahoma, Tennessee, and West Virginia—high-stakes testing was identified by AB as having been adopted prior to 1990 or in 2000. Because these adoptions fall outside of the relevant testing period, any pre/post comparison based on NAEP data is impossible. Thus, we refer to their design as "pseudo-trend" because they frequently lack data before or

Table 3. Changes in NAEP mathematics exclusion rates, by Amrein-Berliner (AB) high-stakes states versus other states: 1992–2000

	Change in fourth-grade mathematics exclusion rates		Change in eighth-grade mathematics exclusion rates	
	1992–2000	1996–2000	1992–2000	1996–2000
AB high-stakes states	3.8	1.3	3.4	2.3
Other states	4.1	2.0	2.6	1.9
High-stakes differential	–0.3	–0.7	0.8	0.4
Statistical significance	*p*<.76	*p*<.44	*p*<.40	*p*<.64

SOURCE: Author calculations.

Table 4. Adjusted average gains in NAEP mathematics scores, by Amrein-Berliner (AB) high-stakes states versus other states: 1992–2000

	Change in fourth-grade NAEP mathematics scores		Change in eighth-grade NAEP mathematics scores	
High-stakes advantage	1992–2000	1996–2000	1992–2000	1996–2000
Unadjusted for test exclusions	5.3	1.9	4.8	2.8
Statistical significance	*p*<.001	*p*<.04	*p*<.003	*p*<.02
Adjusted for change in test exclusions	5.2	2.3	3.7	2.5
Statistical significance	*p*<.001	*p*<.02	*p*<.02	*p*<.02

NOTE: Adjusted average gains come from regression of NAEP score changes on exclusion rate changes.

SOURCE: Author calculations.

after the treatment of interest, and they often have just two or three test scores that are not even aligned with the treatment. For some states, they observe only a single test score change, obviously making any pre/post comparison unreliable.

The use of national average changes in NAEP scores as a reference point further confounds the study. Any effect of accountability systems is already captured in the national score change. By 1996, only 10 states had an accountability system in place, so the effect might not excessively affect the average. But by 2000, a majority of states were on board, so their impacts affected the national average change to a much greater degree. Late-adopting states are effectively being compared to other high-stakes states, making it difficult to show relative gains and completely rendering moot the interpretation that any differences reflect the high-stakes treatment. To take a purely hypothetical example, assume that 6 of the high-stakes states gained 20 percent, while the other 20 gained 2 percent each and the no-accountability states made no gains whatsoever—yielding a national average gain of 3 percent. AB's approach would say that accountability had failed: just 6 states beat the national average, while 20 were below the average. In fact, ignoring any complications of exclusions, AB would report this as something like, "Just 23 percent of states posted gains on NAEP higher than the national average after high stakes were introduced." The right approach, of course, would be to compare gains of high-stakes states to those of no-accountability states.

A subtler but important issue arises when the timing of adoption of an accountability system was bracketed by NAEP tests. It is clear that AB did not use a consistent convention. In some cases, it appears that they used the NAEP results from the period immediately prior and immediately following adoption of accountability, but in others, it appears that they used a different time interval, in some cases starting after the accountability systems were adopted. The one consistent choice appears to be reliance on the least flattering results (for high-stakes accountability).

The implications of these nonscientific procedures is best seen within the context of their finding of "harm." Table 5 examines the set of states where AB concluded that fourth-grade NAEP math scores decreased with the introduction of high-stakes testing. For the eight such identified states, we present aggregate information on testing and results. In three of the eight states (New Mexico, Oklahoma, and West Virginia), AB identify the introduction of high-stakes testing as falling outside the testing period (which did not begin until the 1990s). Moreover, no real trend data in math gains are available for Nevada and Oklahoma, where only a single period of test change is observed. During the 1992–96 period when Kentucky, Maryland, and Missouri intro-

Table 5. Data on NAEP fourth-grade mathematics performance in states identified by Amrein-Berliner (AB) as decreasing after the introduction of high-stakes tests: 1992–2000

States where AB declared decreases in NAEP scores	Introduction of high-stakes testing (AB date)	1992–1996	1996–2000	1992–2000
Kentucky	1994	**4.9**[2]	1.0	**5.9**[3]
Maryland	1993	**3.4**[2]	1.6	**5.0**[3]
Missouri	1993	**2.5**[3]	**3.8**[2]	**6.3**[3]
Nevada	1998	**N/A**	**2.7**[3]	**N/A**
New Mexico	**1989**[1]	0.5	0.0	0.6
New York	1999	**4.2**[2]	**3.9**[2]	**8.1**[2]
Oklahoma	**1989**[1]	**N/A**	**N/A**	**4.7**[3]
West Virginia	**1989**[1]	**8.1**[2]	1.5	**9.6**[2]

N/A—NAEP data unavailable for this time period.

[1] No NAEP tests at or before introduction of high-stakes testing.

[2] Change in NAEP scores exceeds the average change in NAEP both for all states and for states not adopting high-stakes testing.

[3] Change in NAEP scores exceeds the average change for states not adopting high-stakes testing.

NOTE: Bold entries highlight evidence concerns discussed in text.

SOURCE: Author calculations.

duced high-stakes testing, two had math gains exceeding the average for all tested states, and one had gains that just exceeded the average for states that did not introduce high-stakes testing.[15] Nevada, which they record as introducing high-stakes testing in 1998, had gains during 1996–2000 that exceeded gains for non-high-stakes states. Over the entire period of 1992–2000, five of the six states for which data are available showed gains that at least exceeded the average for non-high-stakes tests; New York and West Virginia exceeded the average for all states. And this is the group of states that AB identify as being harmed by high-stakes testing! *Not a single state* provides evidence of harm following the introduction of high-stakes testing. When read correctly, if anything, the evidence points to generally higher performance in this group of states.

The final blow to the credibility of AB's results comes at the point of drawing inferences based on their analysis. Regardless of the choice of design, and ignoring the selective use of NAEP scores, we would still expect AB to consider all the available data *as they had constructed it* to draw conclusions. But they did not. First, they eliminated all information about the magnitude of score changes, relying solely on whether scores increased or decreased. Second, they eliminated all the states that they judged to be "unclear," which reduced the final tally to "improved vs. declined" instead of "improved vs. all states that adopted high-stakes."[16] For instance, they recorded positive or negative results on the NAEP fourth-grade math test for just 12 of the 26 states with high-stakes for grades K–8. AB found that fourth-grade math scores increased at a slower rate than the national average in eight of the remaining states (those in table 5), faster in just four. Yet they write this up in a highly misleading fashion, claiming "67 percent of the states posted overall decreases in NAEP math grade 4 performance as compared to the nation after high-stakes tests were implemented." Actually, AB witnessed gains slower than the national average in just 8 of 26 high-stakes states, or 31 percent.

Instead of concluding that the evidence does not support the proposition that high-stakes accountability increases student achievement, it would be more accurate to say that the chosen evidence by AB does not support any inference at all.

Simply applying the underlying approach of AB to all of the data on NAEP achievement completely reverses their conclusions. High-stakes test states on average perform significantly better than non-high-stakes states. For the reasons described previously, we still do not think that these simple comparisons are the best way to analyze this question, but this analysis demonstrates that there is no difference in the broad results from their crude approaches and the preferred analytical approaches we described previously.

The competing evidence on accountability program performance raises a number of disturbing issues.

Not in a Vacuum

The competing evidence on accountability program performance raises a number of disturbing issues. One is how unaware or indifferent the media and many policymakers are to quality differences in the available evidence. The recent publicity surrounding the AB essay highlights the vulnerability of key public policy initiatives to faulty evidence and badly informed reporting.[17] Distinct from other policy fields, reports in education seem to be taken at face value or—worse—on the political orientations of the authors, independent of the rigor of the analysis or the suitability of the inferences that are drawn. While the most obvious example recently concerned the me-

[15] In terms of what periods were looked at by AB, it is difficult to come up with the rule for decisions on NAEP scores that includes both Maryland and Missouri as "decreasing" states.

[16] As described above, the label "unclear" rests on their strong and untested hypothesis about the impact of exclusion rates on scores. Results are unclear whenever the movement in exclusion rates is the same direction as the movement in test scores, regardless of the magnitude of either change.

[17] Most notable among the publicity was a front page article in the New York Times (Winter [2002]). A link to this article currently appears on the home page for the Great Lakes Center for Education Research and Practice: http://www.nytimes.com/2002/12/28/education/28EXAM.html. Other newspapers and professional publications dutifully provided their own reporting of the AB results.

dia, the problem applies as well to many other actors on the education landscape, including the legislative and executive leadership in many states.

The issue of evidence quality is of prime importance when individuals serve a gatekeeper function for disseminating information to the general public. The media acts as a filter to select issues that merit attention and then distills them into a few key points. Decisionmakers in education agencies serve a similar function when they attempt to reflect the effectiveness of the programs they have implemented. Individuals in these positions are trusted, and expected, to go beyond the press release or a superficial examination of a report or analysis by checking the facts, gauging the credibility of the analytic approach, and vetting the results. We would certainly expect this if the topic under investigation were an allegation of fraud or a new breakthrough in power generation. We need similar assurances in education.

Perhaps this disregard is understandable when one considers that the issue of the quality of evidence has only recently been raised among educators themselves. A recent National Research Council panel was convened to assess "scientific principles for education research"—a type of inquiry unheard of in other research and policy fields (Shavelson and Towne 2002). Most schools of education offer courses in research methods as part of the curriculum, but a wide variety of techniques are taught, determined in no small part by the training and interests of the faculty teaching the courses and not limited to traditional scientific inquiry. This is not to say that there are no appropriate uses for the variety of analytic skills that are taught. However, when significant public policies involving many millions of dollars are on the line, as in the case of school and student accountability programs, evidence must meet the highest scientific standards. The analysis should be rigorous enough to consider and objectively to control for potentially competing explanatory factors, and the resulting evidence must be reliable. It should not be argued on political grounds masquerading as science.[18]

When significant public policies involving many millions of dollars are on the line, evidence must meet the highest scientific standards.

Federal policy has also taken a turn towards more stringent standards of evidence. The No Child Left Behind Act includes strong requirements for employing educational programs based on solid scientific research. The creation of the Institute of Education Sciences is clearly directed at improving the quality of educational research. And, reacting to obvious quality concerns about research that was being used to support policy, the U.S. Department of Education in 2002 funded the What Works Clearinghouse to establish strict scientific criteria for studies on program performance. In an effort to provide a "trusted source of scientific evidence," the Clearinghouse is designed to concentrate primarily on the quality of the research design and the rigor of the analytic techniques. (See http://w-w-c.org)

Reporters should not be expected to be experts in statistical analysis any more than they are expected to be fully versed in biochemistry or investment banking regulations. But it is not unreasonable to hold up a standard of reasonable scrutiny (bringing in expertise if needed as is done for medical and scientific reporting).

It is also not as if the issue is unimportant. Improving our educational performance would arguably lead to greater gains for society than any of the medical breakthroughs of the past decade. For example, had there been true educational improvements following *A Nation at Risk*—putting U.S. student achievement on par, say, with that of students in better performing European countries—it has been estimated that the GDP of the United States would have expanded sufficiently by 2002 to pay for all K–12 expenditures.[19]

[18] See the debates about the effectiveness of accountability systems that entered into the 2000 presidential elections; Grissmer et al. (2000), Klein et al. (2000), and Hanushek (2001).

[19] See Hanushek (2003a, 2003b) (http://www.educationnext.org/20032/index.html).

What We Do Not Know

We have suggestive evidence that accountability as implemented in the 1990s has been helpful. It is clear that, for one reason or another, performance has been better in accountability states than in nonaccountability states. We also have evidence that a number of unintended consequences have followed the introduction of accountability. We do not wish to suggest that we yet have anywhere near the amount of reliable evidence that is needed for developing fully satisfactory testing and accountability systems. But this is far different from completely retreating from assessing and reporting schooling outcomes.

The findings leave us short of what we would like to know for policy purposes.[20] We do not understand how best to design accountability systems that can be directly linked to incentive systems. For example, the vast majority of state accountability systems report average performance for each school on various state tests. These are sometimes disaggregated for, say, race and ethnic groups. But, because these average scores are highly dependent on factors outside the control of schools—such as families and friends—it would not be appropriate to base school performance rewards on these unadjusted average scores. Doing that would encourage schools to concentrate more on who is taking the test than on how their scores can be improved. Incentives are best attached to the value-added for which schools and teachers are responsible.

Similarly, uncertainty remains about the best set of tests to measure accomplishment of the learning standards of each state. Concerns about any possible narrowing of the curriculum or inappropriate changes in instructional practice are in large part concerns about the quality of the testing—because the entire intent of the accountability systems is that teachers do in fact teach to a well-designed set of tests that adequately reflect the range of material that students should know.

Federal legislation in the No Child Left Behind Act represents an important starting point in a process to improve the performance of our schools. It established the necessity for regular annual testing of students and the public reporting of results. It also made some guesses about how to build incentives and requirements into the system. The hope (and intent) of the anti-accountability forces is that regular testing and reporting be nipped in the bud. The challenge to everybody is ensuring that we learn about accountability and adjust any current flaws before the anti-accountability forces succeed. Their success would surely leave our children and our nation worse off.

[20] Issues of accountability system design and of incentive aspects of accountability systems are discussed in Hanushek and Raymond (2003b). These analyses also assess the available evidence on various design issues.

References

Amrein, A.L., and Berliner, D.C. (2002). *The Impact of High-Stakes Tests on Student Academic Performance: An Analysis of NAEP Results in States With High-Stakes Tests and ACT, SAT, and AP Test Results in States With High School Graduation Exams.* Tempe, AZ: Educational Policy Research Unit, College of Education, Arizona State University.

Amrein, A.L., and Berliner, D.C. (2003). Does Accountability Work? [Correspondence]. *Education Next*, *3*(4): 8.

Carnoy, M., and Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis, 24*(4): 305–332.

Fletcher, S.H., and Raymond, M.E. (2002, April). *The Future of California's Academic Performance Index.* Report to the California Secretary of Education. Stanford, CA: CREDO, Hoover Institution, Stanford University.

Grissmer, D.W., Flanagan, A., Kawata, J., and Williamson, S. (2000). *Improving Student Achievement: What NAEP State Test Scores Tell Us*. Santa Monica, CA: RAND.

Hanushek, E.A. (2001). Deconstructing RAND. *Education Matters, 1*(1): 65–70.

Hanushek, E.A. (2003a). The Importance of School Quality. In P.E. Peterson (Ed.), *Our Schools and Our Future: Are We Still At Risk?* (pp. 141–173). Stanford, CA: Hoover Institution Press.

Hanushek, E.A. (2003b). Lost Opportunity. *Education Next*, *3*(2): 84–87.

Hanushek, E.A., and Raymond, M.E. (2003a). Improving Educational Quality: How Best to Evaluate Our Schools? In Y. Kodrzycki (Ed.), *Education in the 21st Century: Meeting the Challenges of a Changing World* (pp. 193–224). Boston: Federal Reserve Bank of Boston.

Hanushek, E.A., and Raymond, M.E. (2003b). Lessons About the Design of State Accountability Systems. In P.E. Peterson and M.R. West (Eds.), *No Child Left Behind? The Politics and Practice of Accountability* (127–151). Washington, DC: Brookings.

Hanushek, E.A., Rivkin, S.G., and Taylor, L.L. (1996). Aggregation and the Estimated Effects of School Resources. *Review of Economics and Statistics, 78*(4): 611–627.

Hanushek, E.A., and Somers, J.A. (2001). Schooling, Inequality, and the Impact of Government. In F. Welch (Ed.), *The Causes and Consequences of Increasing Inequality* (pp. 169–199). Chicago: University of Chicago Press.

Klein, S.P., Hamilton, L.S., McCaffery, D.F., and Stecher, B.M. (2000). *What Do Test Scores in Texas Tell Us.* Santa Monica, CA: RAND.

Peterson, P.E., and West, M.R. (Eds.). (2003). *No Child Left Behind? The Politics and Practice of Accountability*. Washington, DC: Brookings.

Shavelson, R.J., and Towne, L. (Eds.) (2002). *Scientific Research in Education*. Washington, DC: National Academy Press.

Winter, G. (2002, December 28). Make-or-Break Exams Grow, but Big Study Doubts Value. *New York Times*, p. A1.