

Federal Reserve Bank of New York

Economic Policy Review

March 1998

Volume 4 Number 1

Proceedings of a Conference on Excellence in Education

1	To Our Readers	<i>William J. McDonough</i>
7	Creating a Performance-Driven System	<i>Rudy Crew</i>
11	Conclusions and Controversies about the Effectiveness of School Resources	<i>Eric A. Hanushek</i>
29	Reassessing the View That American Schools Are Broken	<i>Alan B. Krueger</i>
47	What Do America's "Traditional" Forms of School Choice Teach Us about School Choice Reforms?	<i>Caroline M. Hoxby</i>
61	Schools and Student Achievement: More Evidence from the Milwaukee Parental Choice Program	<i>Cecilia Elena Rouse</i>
79	What Have We Learned about the Benefits of Private Schooling?	<i>Derek Neal</i>
87	Measuring the Value of Better Schools	<i>Sandra E. Black</i>
97	The Two-Legged Stool: The Neglected Role of Standards in Improving America's Public Schools	<i>Julian R. Betts</i>
117	Standards, Information, and the Demand for Student Achievement	<i>Richard J. Murnane and Frank Levy</i>

EDITOR

PAUL B. BENNETT

ASSOCIATE EDITORS

REBECCA S. DEMSETZ

KENNETH N. KUTTNER

PHILIP E. STRAHAN

JOSEPH S. TRACY

EDITORIAL STAFF

Valerie LaPorte

Mike De Mott

Elizabeth Miranda

PRODUCTION STAFF

Carol Perlmutter

Lingya Dubinsky

Jane Urry

The ECONOMIC POLICY REVIEW is published by the Research and Market Analysis Group of the Federal Reserve Bank of New York. The views expressed in the articles are those of the individual authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

FEDERAL RESERVE BANK OF NEW YORK

**ECONOMIC POLICY REVIEW
ADVISORY BOARD**

Andrew Abel
Wharton, University of Pennsylvania

Ben Bernanke
Princeton University

Timothy Bollerslev
University of Virginia

Charles Calomiris
Columbia University

Richard Clarida
Columbia University

John Cochrane
University of Chicago

Stephen Davis
University of Chicago

Francis X. Diebold
University of Pennsylvania

Franklin Edwards
Columbia University

Henry S. Farber
Princeton University

Mark Flannery
University of Florida, Gainesville

Mark Gertler
New York University

Gary Gorton
Wharton, University of Pennsylvania

James D. Hamilton
University of California, San Diego

Bruce E. Hansen
Boston College

John Heaton
Northwestern University

Richard J. Herring
Wharton, University of Pennsylvania

Robert J. Hodrick
Columbia University

R. Glenn Hubbard
Columbia University

Christopher M. James
University of Florida, Gainesville

Kose John
New York University

Edward Kane
Boston College

Deborah Lucas
Northwestern University

Richard Lyons
University of California, Berkeley

Frederic S. Mishkin
Columbia University

Raghuram G. Rajan
University of Chicago

Kenneth Rogoff
Princeton University

Christopher Sims
Yale University

Kenneth D. West
University of Wisconsin

Stephen Zeldes
Columbia University

Federal Reserve Bank of New York Economic Policy Review

Table of Contents

March 1998

Volume 4 Number 1

Excellence in Education: Views on Improving American Education

Proceedings of a Conference Held at the Federal Reserve Bank of New York in November 1997

- 1 TO OUR READERS
William J. McDonough
- 3 SUMMARY OF OBSERVATIONS AND RECOMMENDATIONS
Joseph S. Tracy and Barbara L. Walter
- 7 CREATING A PERFORMANCE-DRIVEN SYSTEM
Rudy Crew

SESSION 1: EDUCATIONAL RESOURCES AND OUTCOMES

- 11 CONCLUSIONS AND CONTROVERSIES ABOUT THE EFFECTIVENESS OF SCHOOL RESOURCES
Eric A. Hanushek
- 29 REASSESSING THE VIEW THAT AMERICAN SCHOOLS ARE BROKEN
Alan B. Krueger
- 45 SUMMARY OF FLOOR DISCUSSION
Carol Rapaport

SESSION 2: COMPETITION AND CHOICE IN EDUCATION

- 47 WHAT DO AMERICA'S "TRADITIONAL" FORMS OF SCHOOL CHOICE TEACH US ABOUT SCHOOL CHOICE REFORMS?
Caroline M. Hoxby
- 61 SCHOOLS AND STUDENT ACHIEVEMENT: MORE EVIDENCE FROM THE MILWAUKEE PARENTAL CHOICE PROGRAM
Cecilia Elena Rouse
- 77 SUMMARY OF FLOOR DISCUSSION
David Brauer

SESSION 3: PUBLIC AND PRIVATE SCHOOL COMPARISONS

- 79 WHAT HAVE WE LEARNED ABOUT THE BENEFITS OF PRIVATE SCHOOLING?
Derek Neal
- 87 MEASURING THE VALUE OF BETTER SCHOOLS
Sandra E. Black
- 95 SUMMARY OF FLOOR DISCUSSION
James Orr

SESSION 4: THE ROLE OF EDUCATIONAL STANDARDS

- 97 THE TWO-LEGGED STOOL: THE NEGLECTED ROLE OF EDUCATIONAL STANDARDS IN IMPROVING AMERICA'S PUBLIC SCHOOLS
Julian R. Betts
- 117 STANDARDS, INFORMATION, AND THE DEMAND FOR STUDENT ACHIEVEMENT
Richard J. Murnane and Frank Levy
- 125 SUMMARY OF FLOOR DISCUSSION
Margaret M. McConnell

ROUNDTABLE DISCUSSION

- 127 SUMMARY
Joseph S. Tracy and Barbara L. Walter

To Our Readers:

I am pleased to introduce this special issue of the Economic Policy Review, which is devoted to the proceedings of the conference "Excellence in Education: Views on Improving American Education." The conference, held at the Federal Reserve Bank of New York on November 14, 1997, brought together leading economists and educators to explore the best policies for improving the performance of our schools.

All of you know how critical the education of our children is to the future of this country. We must give our young people the knowledge and skills they need to build productive careers and to contribute to a strong economy. As a nation, we want to be positioned to succeed in the global economy, with low levels of unemployment and high levels of expertise, creativity, and initiative.

In the field of applied microeconomics, economists seek to explain what factors make workers more productive. Similarly, educators seek to identify ways of enabling students to learn more effectively. The Fed conference provided a forum in which these two disciplines could share the latest research findings and discuss the strategies that are most likely to strengthen student performance. The conference also benefited from the participation of other key players in the effort to improve education, including representatives of parent organizations and nonprofit groups that fund school reform initiatives.

My hope is that this issue of the Economic Policy Review will spur your thinking about these matters. The volume opens with a brief overview of the themes and recommendations put forward in the papers presented at the conference. The overview is followed by remarks given before the conference by Rudy Crew, chancellor of the New York City public schools, on the creation of a standards-based school system. The papers themselves cover a broad range of topics, including the effectiveness of current school spending, the impact of school choice on student achievement, and the importance of educational standards.

As with most complex issues, there are no simple answers to the question of how we help our schools produce the best educated workers. But the field of economics can help to clarify which policies will have the greatest impact. I am confident that by reading this volume, you will gain a fuller understanding of how we should spend our educational resources to close the gap between the skills we provide our children now and the skills they will need in the increasingly competitive economy of the future.



*William J. McDonough, President
Federal Reserve Bank of New York*

Summary of Observations and Recommendations

Joseph S. Tracy and Barbara L. Walter

The conference “Excellence in Education” focused on a critical question: How can we most effectively improve elementary and secondary education in the United States? Since the early 1980s, the demand for high-skilled workers in the American economy has increased sharply. Yet many high-school graduates are finding that they lack the training to qualify for the types of jobs that would help them secure a traditional middle-class life. It is not that students today are less able to learn; as Richard Murnane and Frank Levy point out in their paper, average math and reading scores are higher now than they were in the early 1970s. Rather, job requirements have been rising quickly while corresponding advances in education have been slow to materialize.

This mismatch is particularly troubling because it has contributed to the widening of income disparities in this country. Individuals at the lowest economic levels are often the least well served by our schools. Thus, they are at a particular disadvantage in competing for jobs. As the skill levels required in the labor market rise, this group is likely to slip further into poverty. Alluding to the income gap in his opening remarks at the conference, William McDonough stressed that “improving education for everyone is the only way to make progress on this problem.”

Because the demands of our economy are changing—and because earlier reforms have not kept pace—many of the conference participants spoke of the need for new education policies. Both Eric Hanushek and Julian Betts remarked in their studies that the United

States has traditionally relied on the quantity of education its citizens receive—that is, the relatively high number of years of required schooling—to be competitive globally. But many countries whose students outperform U.S. students on standardized tests are now beginning to rival the United States in average years of schooling, creating a new and more intense form of competition.

The recommendations put forward by conference participants for dealing with the growing crisis in education fall within several broad groupings, each discussed in more detail below: greater competition, increased choice for parents, stricter accountability for teachers and administrators, the linking of incentives to performance, significant emphasis on the establishment of standards, smaller class size, and more experimentation with a broad range of policy initiatives. With revenues in federal, state, and city coffers growing, politicians and citizens are acquiring additional flexibility to fund education initiatives, so the discussion of policy alternatives in this volume is especially timely. One often hears that making choices is more difficult in good economic times than in bad. But policymakers who read this volume will find much thoughtful analysis to guide their decisions on the best course for the nation’s educational system.

GREATER COMPETITION

Many conference participants identified increased competition among schools and school systems as a key component of any program of reform. In his account of the changes

under way in the New York City public schools, Rudy Crew emphasized that the ability to compete is essential to the survival of public school systems:

The transformation of [the New York City public school] system is driven by the same market forces that drive our economy—namely competition, quality, and productivity. Not only must our students be able to compete in a global, information-age marketplace, but our schools must be able to compete with private and parochial schools as well as the privatization movement. Parents need to know that the product of our schools will be of consistently high quality and that they can count on strong positive outcomes. . . . Ultimately, our schools must perform at a level that restores the public trust in their capacity to fulfill their mission, or we will lose the franchise.

Although alternatives such as charter schools and open enrollment programs are gaining prominence, public schools have long faced competition from other sources. To investigate whether increased competition improves educational outcomes, Caroline Hoxby looked at the traditional forms of school choice in the United States: parents' ability to choose between public and private schools and parents' ability to choose among public school districts by deciding where to reside. Hoxby's analysis suggested that public schools do in fact react to competition by upgrading the schooling they offer and that parents exercising greater choice prompt schools to adopt more demanding curricula and more structured classroom environments. Other conference participants noted an additional benefit of greater competition: schools concerned about enrollments may be more motivated to take on the risks of large-scale, meaningful reform.

The competitive model of education favored by many of the participants entails increased accountability. Under this model, failed schools would be closed quickly and poor teachers and administrators dismissed—albeit with appropriate due process. But this same model of education also recognizes the importance of positive incentives—that is, “carrots” in addition to “sticks.” Voters can be expected to approve more funding for successful

schools, and parents will very likely seek to have their children admitted to such schools, but policymakers must find additional means of rewarding superior teachers, principals, and programs.

INCREASED CHOICE

As Hoxby's paper suggested, the beneficial effects of competition come into play when parents can choose their children's schools. Parents and students alike clearly want more options. In the roundtable discussion that closed the conference, Peter Flanigan noted that his organization received 23,000 applications from public school students for 1,000 scholarships to private schools. Although middle- and upper-income families even now have some freedom to decide which schools their children will attend, low-income families typically have many fewer alternatives.

Increased choice can take different policy forms: choice between public and private schools (aided by vouchers), choice between charter schools and traditional public schools, or choice among traditional public schools. Derek Neal's paper supported the broadly held view that Catholic schools generally provide inner-city minority students with greater skills and higher graduation rates than those offered by public schools. In considering why private Catholic schools achieve a better outcome, Neal emphasized the poor quality of the public school alternatives available to urban minorities. Other conference participants cited parental commitment to education (parents of children attending private school pay tuition) and the private schools' ability to turn away students. Still others attributed the difference in outcomes to features of the private school environment such as reduced bureaucracy, the increased autonomy given to principals, and students' greater sense of personal safety in the classroom. Cecilia Rouse's examination of the Milwaukee Parental Choice Program, a publicly funded program that provides vouchers to low-income students to attend nonsectarian private schools, offered some corroboration of the benefits of private schooling: although the program had no discernible effect on the participating students' reading achievement, the students showed significant gains in math.

BETTER INCENTIVES

All economists are trained in the theory that incentives matter, and they would expect this theory to hold in the schools as in other settings. Yet looking at the broad spectrum of practices and policies in our educational system, conference participants argued that much more can be done to motivate school personnel to improve their programs. The link between teacher pay and student learning was one area identified as requiring significant change, especially since teacher wages tend to be relatively homogeneous and are based more on years of experience and education than on classroom outcomes. Participants favored giving administrators, particularly principals, more incentive tools and greater flexibility in using these tools. Although the support for qualitative performance incentives was strong, participants recognized that further study of the design and implications of such incentives is needed.

EXPLICIT STANDARDS

Several conference participants emphasized the central role of educational standards in the reform of the school system. Betts called for curriculum standards that would clearly delineate what students are expected to learn. He also recommended that students be evaluated regularly to determine whether they are meeting the standards. Among Betts' specific proposals was the adoption of exit exams—a test of basic skills that all students would be required to pass before they graduated from, or dropped out of, high school.

Murnane and Levy approached the subject of standards from a somewhat different perspective. They argued that parents and others lack the information that would allow them to compare the skills that students learn in school with those that are valued in the national economy—namely, superior math and reading skills and the ability to solve problems, to communicate effectively, and to work in teams. The authors suggested that the way to address this lack of information is to establish more rigorous standards, test student performance against those standards, and inform parents of the results. Parents would be much more likely to become involved in the improvement of the schools if they

were presented with evidence that their children were not being prepared to meet the demands of the workplace.

Not all participants agreed with Betts, Murnane, and Levy that a program of standards and testing is the key to better schools. Some expressed the view that teachers will waste time “teaching to the test.” To the extent that standardized tests rely only on rote memorization, this argument has some merit. But if the tests are well designed—for example, if they require written responses that assess students' ability to synthesize information and apply concepts—then teaching to the test is exactly what teachers should do.

One controversial question discussed by participants was whether standards should be established at the national or the state level. There was no consensus on this issue. Although national standards might be more cost effective, participants were quick to point out the difficulty of reaching agreement on a set of national standards. Standards that received the support of educators and policymakers in all states might be too weak to be meaningful.

SMALLER CLASS SIZE

Conference participants expressed different views on the relationship between class size and learning outcomes. Hanushek pointed out that class size has fallen over the past couple of decades, while average scores on international math and reading tests have improved little. Yet Alan Krueger argued that the Tennessee Student-Teacher Achievement Ratio (STAR) experiment, currently the most closely watched state program, did show gains for students in smaller classes. In the first year that STAR students were assigned to smaller classes, they performed better on standardized tests than students assigned to regular classes. Moreover, the beneficial effects seemed to be the greatest for children of poor parents. Although the effect of additional years in smaller classes was more muted, the improvements of the first year remained. Still, the mix of results obtained by researchers suggests that smaller class size alone is not the answer and that other variables—including teacher competency and the enforcement of standards—may influence outcomes.

MORE EXPERIMENTATION AND EVALUATION

A recurrent theme in the conference sessions was the need for more experimentation and follow-up analysis to determine which policies are the most effective in improving educational outcomes. To date, studies of various policy prescriptions have not provided a wholly satisfactory explanation of why some schools teach better than others. As Betts noted in his paper, for example, the establishment of higher expectations for students—whether in the form of more rigorous curriculum standards, increased homework, or stricter graduation requirements—appears to spur student achievement, but the empirical evidence to support this conclusion is limited.

Conference participants also addressed the issue of how we measure the value of educational quality. Attempts to quantify the value of improved education through its impact on wages earned by students later in life have produced mixed results. Sandra Black adopted an alternative approach in her paper: she calculated what people are willing to pay to reside in a community that would allow their children to attend better schools. By determining how an increase in the average test scores of a school affected the price of houses in that school's attendance district, she was able to attach a dollar value to the benefit of higher test scores. This benefit, she argued, can be compared with the cost of an educational program to assess the program's cost-effectiveness.

CONCLUSION

Conference participants took different positions on the nature and extent of the problems affecting American schools. There was considerable agreement, however, on some broad issues. First, most participants saw a need to clarify the goals of elementary and secondary education. If we establish explicit goals for our schools, then we can measure student progress toward the goals and hold our schools accountable for students' success or failure in reaching them.

Second, participants agreed that the problems of the current system are most chronic in urban schools serving low-income students. They also agreed that in the case of these students, we have sufficiently strong empirical evidence to conclude that providing greater school choice could lead to better educational outcomes.

Finally, a consensus emerged at the conference that more studies are needed before we proceed with large-scale reforms affecting the school population as a whole. Existing research cannot justify efforts to expand choice programs to a much broader set of students, nor can it support substantial increases in expenditures on education. For the immediate future, the best course appears to be continued experimentation with different reform initiatives. In addition, as several participants pointed out, the studies undertaken must be carefully designed at the outset to permit a comprehensive evaluation of their results.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

Creating a Performance-Driven System

Rudy Crew

I am here today to discuss what the New York City public school system is working hard to become. Although our task is carried out in a complex economic and political landscape, our strategy is clear, allowing me today to focus on the essential elements of our efforts to transform the city's public schools. The transformation of this system is driven by the same market forces that drive our economy—namely, competition, quality, and productivity. Not only must our students be able to compete in a global, information-age marketplace, but our schools must be able to compete with private and parochial schools as well as the privatization movement. Parents need to know that the product of our schools will be of consistently high quality and that they can count on strong positive outcomes. We must produce quality models that foster healthy competition within the system. Ultimately, our schools must perform at a level that restores the public trust in their capacity to fulfill their mission, or we will lose the franchise.

We are in the process of creating the nation's first performance-driven system. Underlying the efforts to transform this city's public schools is my belief that the whole school system can be a unit of change. In fact, it must be. Public schools can no longer afford to create isolated pockets of effectiveness. Results count. The choices that are made must be replicable and of a scale that challenges the entire system to improve practices and produce better outcomes.

Rudy Crew is chancellor of the New York City public schools. He presented these remarks at the conference luncheon.

A PERFORMANCE-DRIVEN SYSTEM

A performance-driven system promotes competition, maintains quality, and emphasizes a consistently high level of achievement. The goal of a performance-driven system is to set clear standards and to align resources, policies, and practices with the support that students need to hit the target. The following management principles are used at all levels of the organization:

- define clear standards;
- articulate educational strategies designed to enable all students to meet the standards;
- align all resources, policies, and practices to carry out strategies;
- track results; and
- use the data to drive continuous improvement and to hold the entire system accountable for school performance.

The application of these management principles pushes new levels of responsibility down through the system, creating a culture of commitment, not control.

The need for standards and a sharp focus on outcomes is self-evident. Without a clear destination, all roads lead to anywhere. Standards must be measurable, usable, and—given time constraints—practical, and they must apply to all students. There must be clear lines of accountability so that people at all levels understand that they will be held responsible for school performance.

Resources, including time, people, and money, must be aligned with the educational goals of the system. Many people have asked me why I focused on third-grade literacy. It is not an artificial benchmark. If a child has not acquired basic literacy skills by the end of third grade, then the system will spend grades 4 through 8 chasing a genie that has been let out of the bottle. Catching up with illiteracy has cost more time and money than the system can afford. This is not just good management theory; this is central to the survival of the organization.

The public school system does not have a monopoly on the education of its children. In this era of vouchers and privatizers, we must track results, and pedagogy must yield to the numbers. The issue of how we measure is contentious, but we must accept the fact that the system has competition and that our strategies must produce their promised outcome.

We must also think differently about the data. For too long, assessment data have been used to label, blame, and judge both teachers and students. With clear standards, we can establish objective performance measures and begin to use assessment data to drive continuous improvement.

There are fundamental issues that will challenge our success. Can teachers teach to higher standards? Do principals have the time and skills to be instructional leaders? Who ought to control the dollars? How does autonomy play its hand? Can we differentiate our responses to schools along a continuum of performance? How do we reduce the variation in school performance? Ultimately our poorest performing schools define us in the market, but there are at least 900 other pictures to be taken and shared. We have a tremendous perception problem to manage when we go to the city and state governments to convince them that their investment in the public schools will yield results.

SOURCES OF VALUE IN THE SYSTEM

In spite of these challenges, we are moving the system. How? By reviewing our expenditures in terms of sources of value. Given this system's increasing levels of accountability, it is time to think differently about the cost of

school reform. We have to stop measuring school reform as a whole cost and begin to look at the added value that each investment creates. Some critical sources of value in our system include the following:

Standards. Standards are the cornerstone of a performance-driven system. Standards drive instruction, planning, and budgeting. In New York City, the investment in new performance standards, aligned with curriculum and assessments, is a critical tool for assessing the relationship between all expenditures and student achievement.

The Arts. The research is overwhelming that the arts are fundamental to literacy. Yet a full generation of young New Yorkers received minimal exposure to music and art. Most never had even one dance or theater class. Many of our elementary schools had no arts teachers. Our students live in one of the world's most exciting cultural centers. With initiatives such as Project ARTS, we are restoring the arts to our schools systemwide and giving students an opportunity to enter into that world with intellectual and creative curiosity. Not only does this investment in the arts give children another path to academic success, it also provides another path into the world of work and economic opportunities.

Technology. As we enter the twenty-first century, virtually every aspect of our lives is affected by technological innovation. If we want our children to look at the future with opportunity in their eyes and currency in their pockets, they must be able to use technology to learn, produce, and create throughout their lives. Technology gives students an economic base as they leave our schools to enter the world of work; it enables students to say, I can do this.

Uniforms. School uniforms are far more than a symbol of school pride and tradition; uniforms send a message to the thousands of children wearing them that schoolwork is serious business. Standards of dress are directly related to standards of performance in the classroom. Dressing appropriately in school eliminates many of the pressures and distractions of the peer culture that can be obstacles to learning. It also helps young people understand some of the standards they will be expected to meet in after-school jobs, summer jobs, and adult careers.

A Longer Day and a Longer Year. The platform on which the school system is asked to negotiate its reforms is far too small. Continuing to squeeze each new initiative and legislative mandate into a 180-day school year severely limits the quality of implementation and the quality of outcomes. As we ask more of students and teachers, we need to take the constraints off the current system. It will take more than 180 days of schooling to develop a child capable of thriving in the twenty-first century. We need to build flexibility into the school year to foster innovation and creativity in scheduling as we implement higher standards. A system of seamless, year-round education that involves not just more time but strategic uses of that time could help every student achieve at the same high level.

LEADERSHIP

These are the investments we must make if we are to succeed in transforming the New York City public schools into a national model of successful urban education. But it will take more than resources to succeed.

How people lead as we move ahead will be as critical to our success as any other element. There is no blueprint and no one way to move forward. Hope is critical. While skeptics will be welcomed, I have no patience for cynics who bring with them only hopelessness. It will take courage and creativity to face a canvas as large and public as this one and to begin to paint. True leadership will translate a vision into a plan resonating with hope.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

Conclusions and Controversies about the Effectiveness of School Resources

Eric A. Hanushek

Both the U.S. public and U.S. policymakers pursue a love-hate relationship with U.S. schools. While a majority of parents believe that their children's schools are doing well, a majority also believe that the system as a whole needs help. Complicating this view is a variety of concerns about specific aspects of U.S. schools—they are too expensive, too rigid, too elitist, and too unequal.

During the past year, President Clinton has directed considerable government attention to U.S. education. This attention follows the lead of Presidents Bush and Reagan, who also focused on education policy, although the oversight of such policy is not the primary role of the federal government. President Bush, for example, in 1989 convened a historic gathering of the governors of all of the states to focus exclusively on issues of education. The governors set a series of lofty goals for the year 2000, including the goal that U.S. students

should be first in the world in mathematics and science achievement. Unfortunately, we are now close to the year 2000, but we are not close to meeting the set goals.

This paper analyzes the current state of the education system in the United States. In the course of the paper, I will try to point out where controversy exists, particularly in academic discussions.

OVERVIEW

I begin with some overall observations and conclusions. The subsequent discussion will provide some of the relevant evidence and references to support my conclusions.

As a starting point, educational investments are very important to the U.S. economy, a fact that suggests there is much value in an aggressive human capital investment strategy. The U.S. economy has been built up largely by using a skilled labor force and has capitalized on the presence of skills, making human capital investments very important to the success of the overall economy. Moreover, many authors show that the labor market value of the increased skills, as measured by schooling level, has increased dramatically in

Eric A. Hanushek is a professor of economics and public policy and director of the W. Allen Wallis Institute of Political Economy at the University of Rochester.

recent years. I think this valuation demonstrates that the economy continues to need an increasingly skilled labor force. Recent work has also suggested that education is very important in boosting the growth rates of the nation as a whole and that a very important relationship exists between human capital and growth rates. Economists have recently spent considerable time and effort trying to understand why some countries grow faster than others. The majority opinion is that a nation's stock of human capital is an important component of differential growth rates. In addition, we have thought of education as a primary ingredient in providing equal opportunity to all members of society as a way of cutting down or breaking intergenerational correlations of income. Taken together, these benefits provide important and relatively uncontroversial reasons for us to continue our attention to education.

The controversies relate in small measure to how well we have been doing in providing education, but they relate more to what we should do in the future. My way of framing the issues follows.

First, U.S. students do not perform well compared with students from other countries. In international math and science exams, U.S. students have never performed very well relative to students of other countries. To compensate for this relatively low quality, the United States has historically had high levels of school attainment (years of schooling)—that is, the United States has substituted quantity for quality. Now, however, many countries that have had higher student achievement are beginning to rival the United States on quantity grounds. This suggests that the U.S. economy faces new and different levels of competition in the years ahead.

Second, the United States has made steady and large investments in human capital. The resources invested, however, have had little payoff in terms of student performance. Thus, if the United States is to be more competitive internationally in terms of student achievement, some substantially different policies will be required in the future.

Third, the most likely changes required in schools involve radically different incentives for students and for school personnel. Few direct incentives exist today for improved student achievement, and marginal changes in resources or programs are unlikely to have a noticeable effect on overall student achievement.

Fourth, improved education policies will require better measurement of student performance. In addition, such policies will probably require a period of more extensive experimentation with alternative approaches and incentive schemes.

These conclusions are roughly ordered in terms of the amount of evidence and analysis that we have on them and in terms of the amount of consensus or controversy that exists. Regardless of one's views on the underlying controversies, these conclusions indicate to me that the education sector deserves considerable attention. At the same time, the form of this attention is important. Some people have argued that the high rates of return commonly observed for individual schooling clearly justify governmental action. But the case for governmental involvement in education, as opposed to purely private decision making on schooling, requires more than that. Governmental intervention is frequently justified on the basis of external benefits, benefits that go beyond an individual's investment in schooling. Are there external benefits to investing in education? Education is often thought to be a "large externality" undertaking, but identification and measurement of those externalities have proved difficult.¹ My candidate for the most important potential external benefit from investing in education in the United States—which is new in most thinking—is the overall effect on growth rates and the potential to affect the economy. The work supporting this contention is not as refined as you might like: it does not give precise answers, and there are several qualifications. Nonetheless, I think that growth effects are likely to prove to be a very important policy issue.² At the same time, while establishing a role for government research, this issue does not spell out what such a role should be.

U.S. STUDENT PERFORMANCE

It is useful to begin with the performance of students in the U.S. educational system. In doing so, it is natural to contrast performance in elementary and secondary education with that in higher education. I begin with elementary and secondary education. In terms of quality of learning, U.S. schools are not now, and have never been, very competitive when judged by the performance of elementary and secondary schools around the world. Chart 1, drawn from Hanushek and Kim (1996), pre-

Chart 1

International Test Score Performance, by Test Year



Source: Hanushek and Kim (1996).

sents what we know about all international testing of math and science scores for U.S. students.

International examinations in mathematics and science have been given periodically since the 1960s. The examinations have been taken on a voluntary basis by a variable set of countries. While there was some concern about selective test taking in some countries in the early years, that concern has lessened considerably in the later years of testing. Further, Hanushek and Kim show that these tests have considerable validity in describing the quality of a country's labor force. For the analysis here, all the test scores for students in a given country in a given year are combined to produce a single country test score. The scores are placed on a scale where the world mean for each testing year is fifty.

In Chart 1, the year of testing appears along the top of the chart. Normalized scores are given on the vertical axis, making it possible to compare countries over time.

The U.S. performance moves around over time. This drift closely mirrors the average performance of U.S. seventeen-year-olds on the mathematics and science tests of the National Assessment of Educational Progress (NAEP) (see discussion below). Moreover, the key aspect of this figure is that the United States almost always falls below the median of whatever group of countries is taking the test.

The results released in the fall of 1996 for the Third International Math and Science Test placed U.S. eighth graders in the middle of world performance for 1994-95.³ This performance, which is not included in the figure, comes even though a very wide range of forty-one countries participated in the testing. Thus, there is no real change in the latest scores.

The basic story is that the United States has not been doing particularly well in international comparisons. This result is a bit surprising, given that the United States has an economy built on a skilled labor force. You might ask, "How could that be?" While the United States is not doing well, it is producing skilled goods that one might argue require a skilled labor force.

The answer seems to be that over a long period of time, quantity of schooling has substituted for quality. Historically, the United States has had a labor force with more years of

schooling, on average, than the labor forces of other countries, even if these years of schooling have been of lower quality.

That quantitative superiority is ending. Table 1 compares the percentage of students in different countries that have received upper secondary school education, essentially a high school education. These completion rates are broken down by age.

The important part of breaking these figures down by age is that they can be read as the schooling policies of countries in different years. Individuals who are twenty-five to thirty-four years old in 1992 were educated sometime in the 1980s. People aged thirty-five to forty-four were educated in the 1970s. The next group in the table was educated in the 1960s. And the final group went to school in the 1950s.

If we look at the 1980s, it is clear that a large number of countries are rivaling the United States, where 87 percent of students complete their high school education. Three other countries in the Group of Seven have completion rates exceeding 80 percent. Of the countries outside the

Table 1
PERCENTAGE OF POPULATION ATTAINING UPPER
SECONDARY EDUCATION OR MORE
By Country, 1992

Country	Age Group			
	25-34	35-44	45-54	55-64
GROUP OF SEVEN				
Canada	81	78	66	49
France	67	57	47	29
Germany	89	87	81	69
Italy	42	35	21	12
United Kingdom	81	71	62	51
United States	87	88	83	73
OTHER				
Australia	57	56	51	42
Austria	79	71	65	50
Belgium	60	52	38	24
Czechoslovakia	87	79	68	51
Denmark	67	61	58	45
Finland	82	69	52	31
Ireland	56	44	35	25
Netherlands	68	61	52	42
New Zealand	60	58	55	49
Norway	88	83	75	61
Portugal	21	17	10	7
Spain	41	24	14	8
Sweden	83	76	65	48
Switzerland	87	84	78	70
Turkey	21	14	9	5

Source: U.S. Department of Education (1996b).

Group of Seven listed in the table, another five have completion rates above 80 percent. These numbers contrast sharply with those in earlier decades, when the United States had a very dramatic lead in terms of quantity of schooling. Clearly, other OECD countries and developing countries have dramatically increased the amount of schooling their youth receive. The United States' advantage in quantity of schooling is quickly disappearing.

Charts 2 and 3 provide pictures of science and mathematics achievement in the United States as measured by the NAEP, which is currently the best yardstick of student performance. The heavy line reflects the average scores of seventeen-year-olds on the NAEP over time. What we see from Chart 2 is that today our students are not doing quite as well in science as they did in 1970 (even though, as described below, we have been increasing real per pupil spending steadily over this period). Chart 3 shows essentially the same thing for math, except that instead of declining, 1996 performance is at about the same level as in 1970. This picture does not lead anyone to believe that our investment policy is soon going to address the quality concerns and to push us to the top of the international rankings. "First in the world in math and science in the year 2000" was the goal set forth by the 1989 National Governors' Conference. It does not look like we are on that path.

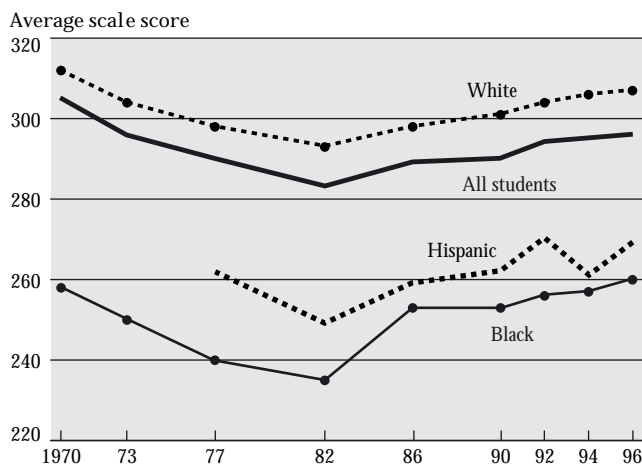
Charts 2 and 3 also suggest that there is a substantial gap between whites on the one hand, and blacks and Hispanics on the other. The gap has narrowed some, but it remains substantial and may have even widened in the most recent period. This disparity goes back to the equality of opportunity concerns; it is also consistent with several analyses that identify the importance of student achievement in explaining some of the college attendance gaps across different segments of society. Those attendance gaps exist throughout this period and seem related to quality of schooling.

The situation with higher education is very different. U.S. higher education is arguably the best in the world. Admittedly, data on higher education are not nearly as good as the data on elementary and secondary education. It is particularly hard to document quality because we do not have good, objective measures. Here is what we do know:

- U.S. business and industry are now willing to pay much more for college graduates than they were in the past, both in relative and in absolute terms;
- foreign students like to come to U.S. higher education institutions, while they do not seem to want to come to U.S. elementary and secondary schools; and
- employers seem much more pleased, at least in their public testimonials, with higher education than they are with elementary and secondary education.

Chart 2

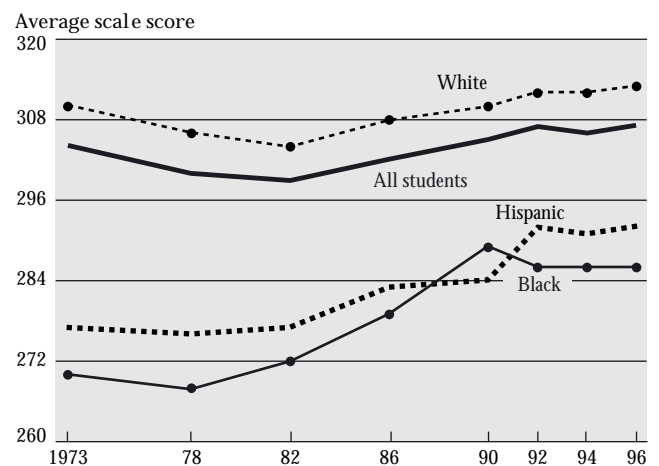
Science Achievement
Seventeen-Year-Olds, by Race/Ethnicity



Note: Science achievement is measured by performance on the National Assessment of Educational Progress (NAEP) exam.

Chart 3

Mathematics Achievement
Seventeen-Year-Olds, by Race/Ethnicity



Note: Mathematics achievement is measured by performance on the National Assessment of Educational Progress (NAEP) exam.

This adds up to a *prima facie* case that quality does not appear to be a major problem in higher education. By contrast, the data displayed before suggest that quality is the major concern in elementary and secondary education. The contrasting picture makes the recent concentration in 1997 on higher education by the President and the Congress puzzling—at least if the policy initiatives are viewed in terms of education as opposed to pure distributional politics. Perhaps the one rationalization is that the call for expanding access to schools—“making the fourteenth year the norm”—is just an extension of the historic policy of substituting quantity for quality. Without pursuing the issues of higher education, I simply assert that elementary and secondary school issues are the most important and pressing. Thus, the remainder of this discussion concentrates exclusively on these issues.

A BRIEF HISTORY OF SPENDING GROWTH

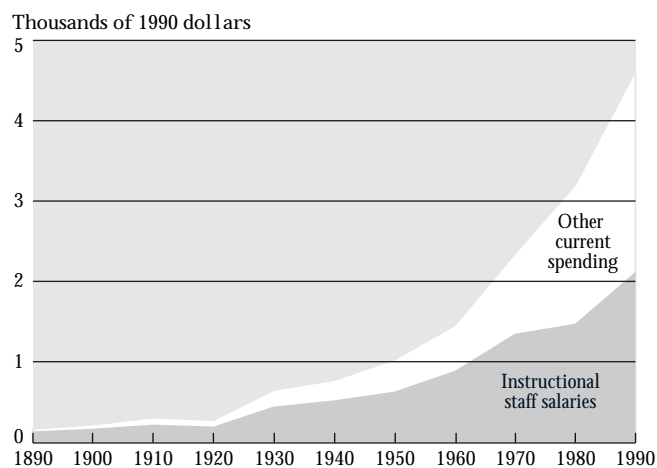
The United States has had a consistent focus on education over a long period of time. This fact surprises many people in the United States. Statements about “how important it is that President Clinton has recently focused attention on education” are common. Implicit or explicit in such discussions is the sentiment that we have been shortchanging the educational system. It may be that the President can get the attention of the population better than anybody else, but a steady policy thrust and heavy weight have been given to education and human capital investment for a long time. This focus on education, however, has not been at the federal government level.⁴ Taking the long view, between 1890 and 1990, we note that real public expenditure on primary and secondary education in the United States rose from \$2 billion to more than \$187 billion.⁵ Significantly, this almost hundredfold increase is more than triple the growth rate of GNP during the same period: current educational expenditure increased from less than 1 percent of GNP in 1890 to 3.4 percent of GNP in 1990.⁶

While increasing enrollment accounts for a portion of the rise in spending, the rise in per student expenditure explains the bulk of the change in educational outlays (Chart 4). Real per student expenditure roughly quintupled in each fifty-year period between 1890 and 1980: it went from \$164 in 1890 to \$772 in 1940, and to \$4,622 in

1990. If we divide per student expenditure into salaries for instructional staff (teachers and principals) and all other expenditure, we find that the unmistakable pattern here is the relative growth of expenditure outside of instructional staff salaries: such spending went from 25 percent of total current expenditure in 1890 to 33 percent in 1940, and to 54 percent in 1990.

Two factors stand out as being of primary importance in explaining total instructional salary spending over the entire 100-year period: the rising price of instructional staff and the declining pupil-staff ratio. Rising teacher salaries were clearly a consequence of economywide labor productivity growth, although the extent to which teacher salaries changed relative to those of other workers is an important issue. By contrast, the decisions leading to reductions in the pupil-staff ratio despite the rise in teacher costs suggest a long-term policy of attempting to raise school quality by reducing the pupil-teacher ratio.⁷ There is substantial debate over the extent to which external changes, notably the expansion of special education, contributed to the decline in the pupil-teacher ratio during the 1970s and 1980s. The analysis by Hanushek and Rivkin (1997) indicates that special education has been important but is still not the largest influence. The growth in special education over the 1980s may have accounted for one-

Chart 4
Spending per Pupil, 1890-1990



Source: Hanushek and Rivkin (1997).

fifth of the growth in spending. (Yet, because of the smaller overall spending growth in the 1990s, this percentage has almost certainly gone up.)

PRICE AND QUALITY OF TEACHERS

The desire to reduce classroom sizes has increased the demand for teachers. At the same time, rising labor market wages for both college-educated men and, particularly, college-educated women have reduced teacher supply. These supply and demand movements offer a straightforward explanation for the teacher price increase from \$34 per day in 1890 to more than \$177 per day in 1990, an increase that accounts for more than 40 percent of the increase in total expenditure on instructional staff over the century. But these numbers tell only part of the story. Schools have also been able to adjust the average quality of teaching personnel by shifting teacher salaries to accommodate shifts in supply and demand. Since potential teachers clearly differ in their skills and consequently their alternative wage opportunities, any increase or decrease in teacher wages beyond that occurring in other sectors reflects a change in where teachers are drawn from the distribution of workers. This fact would be expected to influence teacher quality in the long run.⁸

To trace teacher quality changes, I use annual earnings data for teachers from the six decennial censuses of population taken between 1940 and 1990.⁹ Teacher earnings are compared with the earnings of those who do not teach. Specifically, our primary measure of potential teacher quality is the location of average teacher earnings in the distribution of nonteacher earnings.¹⁰ The lower the percentage of nonteachers who earn less than the average teacher, the worse the teaching jobs when compared with alternative occupations. The use of percentile rankings as opposed to a comparison of mean earnings reduces problems associated with the census' top-coding of incomes and lessens the impact of changes in the tails of the nonteacher earnings distribution.

The movements in relative earnings of teachers have been dramatic. As shown in Table 2, however, they differ noticeably for men and women. While the average male teacher earned more than 84 percent of all males in 1940,

Table 2
RELATIVE SALARIES OF U.S. TEACHERS
By Gender, 1940-90

Year	Percentage of Male Nonteachers Earning Less Than Average Male Teacher		Percentage of Female Nonteachers Earning Less Than Average Female Teacher	
	All Workers	College Graduates	All Workers	College Graduates
1940	84.0	52.5	92.3	68.7
1950	73.4	36.2	86.7	55.0
1960	63.3	28.7	86.9	52.7
1970	62.2	25.7	85.8	47.1
1980	53.0	31.0	77.7	50.1
1990	64.0	36.5	75.1	45.3

Source: U.S. decennial census of population, public use microdata, 1940-90.

this figure fell to 64 percent by 1990. All of this relative fall, however, occurred before 1960; following a slight dip in the 1970s, male teachers have been moving up the earnings distribution. The overall decline in the relative position of women teachers has been almost as large, although female teachers are still better positioned in the earnings distribution than male teachers. The time path of the decline for females has been very different, however, with the largest declines occurring after 1970, when the average teacher moved a full 10 percentage points down the earnings distribution.

The implication of this finding is that schools, while spending increasing amounts on teachers, have also tended to let the quality of teachers slip. This story appears to be closely related to arguments such as those of Baumol (1967). He suggests that a sector subject to low rates of productivity improvement, frequently for technological reasons, will find its wage bill and costs rising relative to those of more advanced sectors. While there are reasons to be skeptical about the magnitude of any such effects (Hanushek 1997), the increasing labor costs and potentially declining teacher quality are consistent with this argument.

RECENT CHANGES IN SCHOOL RESOURCES

The search for more experienced and better educated teachers has been one of the enduring policy thrusts of the last half century. This thrust, along with the desire to reduce class sizes, has been a dominant component of aggregate changes in schools.

Table 3 tracks these changes from 1960 to 1991. Teacher education has increased dramatically, so that more

than half of all teachers had master's degrees in 1991. Moreover, teacher experience, while following some demographic cycles, has reached a very high level. Table 3 also shows the decline in pupil-teacher ratios and the increase in real spending per pupil. Because teacher salaries are closely linked to experience and education, and because variations in salaries and pupil-teacher ratios are the most important determinants in spending per pupil, the added real resources directly drive spending. Between 1960 and 1991, real spending per pupil almost tripled.

It takes little effort to see the contrast between the growing resources in Table 3 and the flat student performance in Charts 2 and 3. At the very least, this contrast suggests the possibility that something is very wrong.

OTHER OBSERVATIONS ABOUT SCHOOL EFFICIENCY

The preceding sections make a strong case that spending and policy toward schools in the United States have not been very well directed. Specifically, spending has improved dramatically, but student performance—at least over the period in which it has been measured—has not improved. This outcome is the very essence of inefficiency: more resources devoted to schools have not improved output.

The aggregate data, however, could be misleading. First, costs not directly related to the typical student—for example, costs for special education—might figure importantly in the rise in spending. Second, the other influences on student performance—families and friends—could have been less favorable over time. For example, some analysts note that single-parent families have increased in recent decades. Also, the percentage of students in families below the poverty level has increased. Factors such

as these could mean that more school resources are needed to overcome existing deficits.

These issues have been debated in considerable detail elsewhere, and there is, in my judgment, a *prima facie* case against them. Nonetheless, it is also worth mentioning the detailed microlevel evidence about the impact of resources. The interpretation of the microlevel evidence has been the most controversial part of the academic debate; these controversies are less relevant to the policy debate.

The investigation of the effects of school resources began in earnest with the publication of the “Coleman Report” (Coleman et al. 1966). This congressionally mandated study by the U.S. Office of Education startled many observers by suggesting that schools did not exert a very powerful influence on student achievement. Subsequent attention was directed at providing additional evidence about the effects of resources.

Over the past thirty years, a steady stream of analyses has built up a consistent picture of the educational process. Studies of educational performance, generally following statistical analyses of the determinants of student achievement, include a variety of different measures of resources devoted to schools. Commonly employed measures include (1) the real resources of the classroom (teacher education, teacher experience, and teacher-pupil ratios); (2) financial aggregates of resources (expenditure per student and teacher salary); and (3) measures of other resources in schools (specific teacher characteristics, administrative inputs, and facilities).

The real resource category receives the bulk of attention for several reasons. First, this category best summarizes variations in resources at the classroom level. Teacher education and teacher experience are the primary determinants of teacher salaries. When combined with

Table 3
PUBLIC SCHOOL RESOURCES IN THE UNITED STATES, 1961-91

Resource	1960-61	1965-66	1970-71	1975-76	1980-81	1985-86	1990-91
Pupil-teacher ratio	25.6	24.1	22.3	20.2	18.8	17.7	17.3
Percentage of teachers with master's degree	23.1	23.2	27.1	37.1	49.3	50.7	52.6
Median years of teacher experience	11	8	8	8	12	15	15
Current expenditure per pupil (1992-93 dollars)	1,903	2,402	3,269	3,864	4,116	4,919	5,582

Source: U.S. Department of Education (1996a).

Note: Per pupil expenditures are based on students' average daily attendance.

teachers per pupil, these variables describe variations in the instructional resources across classrooms. Second, these measures are readily available and well measured. Third, they relate to the largest changes in schools over the past three decades. Table 3 displays the dramatic increase in these school inputs, with pupil-teacher ratios falling steadily, teacher experience increasing, and the percentage of teachers with a master's degree actually doubling between 1960 and 1990. Fourth, studies of growth in performance at the individual classroom level, commonly thought to represent the superior analytical design, frequently have these resource measures, but not the others, available.

These studies yield a simple conclusion, one that is supported in detail elsewhere (Hanushek 1997): there is no strong or consistent relationship between school resources and student performance. In other words, there is little reason to be confident that simply adding more resources to schools as currently constituted will yield performance gains among students. Studies of class size and pupil-teacher ratios, of teacher education, and of teacher experience give little if any support to policies of expanding these resources. This finding has obvious policy implications. Before turning to these, it is useful to clarify precisely what is and is not implied by the data.

Perhaps the most important fact to underscore is that this finding does not imply that all schools and teachers are the same. Quite the contrary. Substantial evidence suggests that there are large differences among teachers and schools. The simple fact remains that these differences are not closely related to teacher salaries or to other measured resources devoted to programs. The Coleman Report, which found that measured school resources explained a small portion of the variance in student achievement, has been commonly interpreted as implying that "schools don't make a difference." This interpretation confuses the effects of measured differences with the full effects of schools and has been shown to be wrong. There is a significant difference between measured resources (of the kind on which policy frequently focuses) and the true effects of schools.¹¹ In fact, it is just this difference between true effects and those of standard resources that forms the basis for the policy considerations below.

The preceding interpretations of the general ineffectiveness of school resource policies has been challenged by some researchers. Two separate challenges deserve attention before I discuss policy implications.

LABOR MARKET OUTCOMES

Taken as a group, the production function studies give little indication that variations of resources have anything to do with variations in student performance. However, the widely publicized findings of Card and Krueger (1992) indicate that variations in school resources are related to earnings differences among workers.¹² The Card and Krueger analysis begins with samples of adult workers from the 1970 and 1980 censuses of population and fills in information about the schooling circumstances of individuals by using information about their year and state of birth. Card and Krueger find that labor market earnings are directly related to school resource differences.

Several factors could contribute to reconciling these conclusions: differences in levels of resources considered by each study, differences in measurement of student performance, differences in specification, and aggregation bias in the statistical analysis.

The workers in Card and Krueger's sample attended schools between the 1920s and the 1970s, a span of time encompassing variations in the level of resources going far beyond what is found today. This suggests one reconciliation of the conflicting study findings: if added resources have diminishing effects on student achievement, current school operations may be largely "on the flat" of the production function, while Card and Krueger observe ranges from the past where resources had stronger effects. A related possibility might be that the political economy of schools has changed over time. For example, with the rise of teachers unions and the resulting change in bargaining positions, resources might be used in different ways and have different student achievement implications now than in the past (see, for example, Borland and Howsen [1992], Peltzman [1993], and Hoxby [1996]). In other words, it is quite possible that the enormous changes in educational resources did have an effect on outcomes in the first half of this century, but that more recent studies are also correct in

finding “no effect” for the sorts of resource changes discussed in current schools.

A series of more technical discussions has also been introduced to “choose” between the competing views of the effects of school resources. The debate has focused on questions about the appropriateness of measuring student performance with achievement tests and on questions about specific aspects of the Card and Krueger statistical methodology. These debates go far beyond this specific paper, but they provide, in the opinion of one of the principals in the debate, a strong foundation for accepting the basic conclusion that added resources have not and are unlikely to improve student outcomes noticeably.¹³

META-ANALYSIS AND THE SUMMARY OF RESULTS

In some research areas, such as the investigation of the health effects of a certain drug therapy, there is frequently an interest in compiling results from a variety of trials. Specialized techniques to combine the results of separate studies and thus assess the magnitude and significance of some relationship have been developed. These approaches go under the general title of “meta-analysis.” The previous summary of results represents one simple approach to the aggregation of results, but other researchers have attempted to do formal statistical tests.

A well-known application of formal statistical tests to education production function data is found in Greenwald, Hedges, and Laine (1996). The question they pose is whether there is any evidence that resources or expenditure differences *ever* appear to affect student performance. Their formal tests lead to rejection of this restricted null hypothesis. The most basic problem with their statistical analysis is that it addresses an uninteresting question from a policy viewpoint. Their results are sometimes interpreted as refuting the conclusion that educational inputs do not affect performance. But in my view, this work both confirms the previous substantive results and points to the same policy conclusions. As all of the analysis shows, productive results are possible, even if seldom achieved currently. This conclusion is central to much of the policy discussion.¹⁴

THE SCHOOLS-ARE-DOING-FINE SCHOOL

Surprisingly, in the face of evidence such as that presented here, some commentators have argued that U.S. schools look pretty good. Krueger (1998) has fallen in line with Berliner and Biddle (1995) and other writers who suggest that concerns about the performance of our education system are quite overblown and that in reality there is evidence both of high performance and of marked improvement. Because these arguments have received wide circulation—largely, it appears, from people who wish to maintain and expand the current structure—it is useful to understand how these arguments are constructed.

Krueger presents evidence about performance on the NAEP exams that he generally interprets as supporting the effectiveness of current schools. For most of his discussion, he combines scores on the reading, mathematics, and science exams for a random sample of students aged nine, thirteen, and seventeen between 1969 and 1996. He places great weight on the occurrence of a statistically significant time trend in scores and a statistically significant correlation between scores and spending per pupil in a majority of the nine tests.

First, note that these tests are not independent of each other. The cohort of students tested at age nine in any year is tested again four years later at age thirteen and again four years later at age seventeen. The same holds true for the cohort of students tested at age thirteen and again four years later at age seventeen. Thus, only two of the nine cohorts of nine-year-old students and two of the nine cohorts of thirteen-year-old students (those tested in 1994 and 1996) are not retested and included in the NAEP results for later ages. For policy purposes, we are clearly most interested in performance at age seventeen, just before students go into the job market or into postsecondary education. The combined analysis by Krueger provides equal weight to test measures at any point during the educational process, even if these measures are “superseceded” by measures closer to the finish of the process and closer to the time when they have real importance.

The importance of combining the tests is immediately apparent from looking at Charts 2 and 3 in this paper and by comparing overall NAEP results on the per-

formance of students around the time that they leave high school with their performance in earlier grades. For science, the average scale score of seventeen-year-olds falls 9 points (0.2 standard deviation) between 1969 and 1996. For math, seventeen-year-olds improve 3 points (0.11 standard deviation) between 1973 and 1996. For reading (not shown in the charts), the seventeen-year-olds improve 2 points (0.05 standard deviation) between 1971 and 1996. (Writing performance, which is only available since 1984, shows a fall of 7 points, or 0.2 standard deviation, by 1996.) Only the fall in science (and in writing since 1984) is a statistically significant difference. By contrast, there are statistically significant increases in five out of six tests of the earlier age groups for science, math, and reading. In other words, our third graders are becoming better at conversation and better at making change, even if they ultimately cannot fill out a job application or participate in modern quality control functions at the workplace. It is these latter findings for early ages that permit Krueger to discuss the “surprising” finding that there is a statistically significant trend in student performance.

Imagine a scoring system for the summer Olympics where the times of runners in the 100-meter dash are recorded at 50 meters, 75 meters, and 100 meters, and where these intermediate and final times are simply averaged to make judgments about who is the fastest sprinter. This is the system Krueger suggests for evaluating the performance of U.S. elementary and secondary schools. It is possible that an Olympic track coach would want to use the information about performance at intermediate distances to assess which aspects of the race each runner should work on. Similarly, one might want to assess whether the U.S. schools are improving or lagging at different grade levels. It is doubtful, however, that the highly aggregated evidence of the NAEP scores would be good evidence for making specific resource allocations. Krueger’s use is solely to support the case that U.S. schools are performing quite well, at least compared with his prior perceptions.

Second, Krueger goes on to use the same basic methodology to assess the effectiveness of school resources. He regresses spending per student in each test year on the existing nine NAEP scores. The entire exercise is peculiar

in the sense that one would not normally expect the performance of seventeen-year-olds to be dictated by just the spending in the current school year. Krueger is relying on the fact that spending has moved up so steadily over the past forty years that the current spending in any year is a good index of differences in the cumulative spending over the school years for different cohorts.

It is no surprise that the results of this statistical exercise for spending are very similar to the previous investigation of trends. Chart 4 and Table 3 show that spending has moved quite steadily upward across the period of the NAEP scores, indicating that either current or cumulative spending will look very similar to a time trend. From analyzing the cumulative average spending for each of the tested cohorts (that is, averaging spending across the prior years of schooling for each age and test year group), the statistical results show the same basic pattern as the simple trends, even if there are minor changes in which estimates are labeled statistically significant. In this version of the trend exercise, neither the negative relationship between the science performance of seventeen-year-olds and spending nor the positive relationship for math performance is statistically significant at the 10 percent level. The positive correlation for reading is statistically significant at the 6 percent level. Again, there is a difference for younger cohorts in each of the tests. For age nine and age thirteen, four out of the six correlations are statistically significant at the 5 percent level, and five out of six are statistically significant at the 10 percent level. The overall results suggest, nonetheless, that good performance in earlier years has not translated effectively into higher cumulative performance at the time of graduation, when achievement counts most.

Third, the translation of the analysis into the relationship between scores and spending allows Krueger to perform a policy analysis that provides some feel for the magnitude of the results. Statistical significance helps to decide whether or not we should believe there is any relationship at all, but it does not indicate how much we might expect from an increase in spending. In order to be concrete, Krueger considers a simple increase in spending of \$2,000 per student. We should first be clear about the meaning of this. With 50 million students in

1995 (45 million in public schools), this specification amounts to an increase in annual spending of \$100 billion. In other words, he is proposing increasing total 1995 spending on elementary and secondary students by more than one-third. If financed entirely through appropriations by the U.S. Department of Education, this illustrative calculation would call for a quintupling of federal aid to schools.

On the basis of the education system's past performance (captured by the simple regressions of NAEP scores for seventeen-year-olds on cumulative prior spending), student achievement would be predicted to rise 0.06 standard deviation in reading and 0.10 standard deviation in math; it would be predicted to fall by 0.08 standard deviation in science. It would seem hard to get much political or public support for this pattern of results, even at substantially lower cost. These projections are obviously not based on the scaling of the charts assessing student performance. Nor does translating these projections into movements across a normal distribution of the population increase the sense that this is what we would call a successful \$100 billion annual expenditure.

Betts (1998), using a different approach, calculates the rate of return to additional school district spending. His findings confirm my conclusion: on the basis of past performance and the current structure of schools, additional spending on schools appears to have a net negative rate of return.

In sum, a variety of researchers and commentators have put a spin on the performance data for U.S. schools that suggests that our schools have been doing well and have been improving as a result of past spending increases. These assertions are not supported by the data. Allowing these arguments to distract us from developing more effective policy options would be a mistake. Contrary to the assertions of Berliner and Biddle (1995), the problems of America's schools represent neither myth nor fraud but instead a series of more fundamental organizational problems.

POLICY IMPLICATIONS

The interpretation of the overall results about school resources depends fundamentally on how the policy and decision-making process is conceived. At one level, these conclusions

clearly imply that educational policymaking is more difficult than many would like. If resources had a consistent and predictable effect on student performance, policymaking would be straightforward. State legislatures could decide how much money to invest in schools and could trust local districts to apply funds in a productive manner. But the fact that local districts do not use funds effectively complicates this picture. The clearest message of existing research is that uniform resource policies will not work as intended.

Similar policy dilemmas face the courts in school finance cases. The courts have entered into education decision making by ruling on suits brought by people who believe that state legislatures are not fulfilling their constitutional obligation to provide equitable or adequate education to particular students in each state. While frequently motivated by concerns about student achievement, in reality both the judicial statement of the issue and the proposed remedies center fundamentally on the level and distribution of resources. If resource availability is not a good index of educational outcomes or if providing for overall resource levels does not ensure a desired level of performance, the courts face the same dilemma as legislatures. Simply providing more funding or a different distribution of funding is unlikely to improve student achievement (even though it may affect the burdens of school financing on the citizens of a state).

A variation of this general theme is to argue that, while resources alone may not be sufficient to guarantee achievement, adequate resources are surely necessary. Undoubtedly, this statement is accurate at some level, because a school with no funds would not be expected to add anything to student achievement. Nonetheless, as shown in Table 3, real spending per student increased by more than 70 percent between 1970 and 1991, even though student performance appears to have been essentially unchanged. Further, nothing in the previous analytical results about the effects of resources suggests that there is a level below which resources have clear and powerful effects on achievement that would be a demonstration that some schools are below the threshold of "necessity." Just asserting that there is some level of necessary expenditure does not make the case for pure resource policies in today's schooling environment. While it is not possible to define scientifically how much is "necessary," the dramatically

larger spending of today has clearly taken almost every school system in the country beyond some minimal level.

This policy conundrum is precisely what led the Panel on the Economics of Education Reform to concentrate not on the specific resources and policies of schools but on the incentive structure. Its report, *Making Schools Work*, emphasizes the need to alter current incentives in schools radically (Hanushek et al. 1994). The simple premise is that the unresponsiveness of performance to resources largely reflects the fact that very little rests on student performance. Because good and bad teachers or good and bad administrators can expect about the same career progression, pay, and other outcomes, the choice of programs, organization, and behaviors is less dependent on student outcomes than on other things that directly affect the actors in schools.

Underlying this view is a more benign opinion of school personnel. Specifically, school personnel are not just ignoring a set of policies that would lead to obvious improvements but instead are simply following existing incentives. An added part of this argument is that the kinds of policies that will work in given situations with given personnel and students vary and that these policies are not easily described and centrally regulated. The assumption is that, given better incentives, school personnel can be motivated to search out what will work in their specific situation. Under current incentives, they appear to devote more of their attention and energies elsewhere.

Earlier work on educational production has provided substantial evidence that vast differences exist among teachers and schools. It is just that these differences are not easily explained by the resources employed or by any simple set of programmatic or behavioral descriptions. The existence of effective teachers and schools, however, implies that one approach to policy is to devise ways of rewarding better performance whenever it is found. In other words, even if the details of what will work are unavailable before the fact (or even after the fact), policy can be described in terms of outcomes, and good outcomes can be rewarded.

Such a description is itself much too simple because we have limited experience with alternative incentive schemes (Hanushek et al. 1994). The alternative incentive structures include a variety of conceptual approaches to providing rewards for improved student performance; they range from merit pay for teachers to charter schools to privatization to vouchers. These are contentious proposals, in part because the introduction of performance incentives might lead to a variety of people other than current school personnel making decisions and even providing educational services. Incentive proposals also could work well or poorly, depending on the details. The purpose here, however, is not to consider the pros and cons of alternatives, but to emphasize the radically different perspective on policy that is embedded in each. Performance incentives recognize that varying approaches by teachers and schools might be productive. Thus, they avoid the centralized “command and control” perspective of much current policy. At the same time, they recognize that simply decentralizing decision making is unlikely to work effectively unless there exist clear objectives and direct accountability.¹⁵

Given the current lack of knowledge about the design or implications of performance incentives, an aggressive program of experimentation and evaluation seems very appropriate (Hanushek et al. 1994). Nonetheless, the lack of direct information should not be taken as support for more of what we are doing now. We actually have considerable experience with the current organization, and current approaches appear to offer little hope for general improvement.

The existing work does not suggest that resources never matter. Nor does it suggest that resources could not matter. It only indicates that the current organization and incentives of schools do little to ensure that any added resources will be used effectively. Faced with this problem, some simply declare that we should still pursue general resource policies but that we should not pursue programs that do not work. This approach would be fine if policymakers could reliably identify programs that do and do not work. We know that their judgments have not been accurate in the past.

ENDNOTES

1. The consideration of externalities is one area in which the current U.S. situation differs from that in many other countries. At the high levels of attainment in the United States, the case for strong influences on literacy, the functioning of democracy, the health of the population, or the reduction in crime—traditional items cited as externalities—seem weak. At lower levels of attainment, a better case might be made. See Hanushek (1996b) and Poterba (1996).
2. While I am perhaps biased toward the work I have done with Kim (1996), the effects of qualitative differences in student performance on national growth rates appear huge. This evidence fits nicely with concerns that have been raised in the United States about the achievement of students—concerns that are developed in the next section.
3. The performance of U.S. eighth graders was relatively better in science than in mathematics. This finding is not completely consistent with the U.S. performance over time in the separate mathematics and science tests of the NAEP.
4. The federal government is not the main actor in either the elementary and secondary or the higher education arena. Of the \$260 billion in revenues for elementary and secondary schools in 1993-94, only 7 percent came from federal funds, while 45 percent came from state funds and the remainder was supplied by local school districts. Similarly, of the \$179 billion in revenues for higher education, slightly more than 12 percent came from the federal government (U.S. Department of Education 1996). More important, resources to schools at all levels have shown considerable growth for a long time, even though most of these funds have come from states and localities.
5. Details of the data and analysis of costs can be found in Hanushek and Rivkin (1997). All monetary measures are adjusted by the GNP deflator to constant 1990 dollars. Moreover, most discussion concentrates on current expenditure, that is, total expenditure less capital investment. Unless otherwise noted, we also use public school expenditure, excluding that going to private schools. The proportion of students attending private schools has ranged from 10 to 15 percent for the entire century, with a vast majority of private school students attending Catholic schools.
6. Spending as a percent of GNP actually peaks around 1975 at 3.9 percent. As noted below, this percentage is affected noticeably by the demographics of the school-age population, making it a poor statistic for comparing the intensity of resources devoted to schooling.
7. An alternative explanation is that schools attempted to protect employment during periods of falling enrollment. While this may have been the initial motivation, subsequent data indicate that the reductions have been permanent ones, resistant to any increase brought about by rising enrollment. The efficacy of improving school quality by reducing pupil-teacher ratios has at the same time been seriously questioned.
8. This is not to say that nonpecuniary factors are unimportant in determining whether individuals choose to teach. Rather, we assume that nonpecuniary benefits or costs of teaching have not changed in comparison with those in other occupations, in which case changes in relative earnings function as a good index of where teachers fall in the labor force. The influence of changes in wage rates on the stock of teachers is complicated, and it depends on the choices of school districts and the behavior of prospective teachers. See Ballou and Podgursky (1995) for a consideration of various outcomes from wage changes.
9. A more complete analysis of spending changes, along with definitions of the precise data used, can be found in Hanushek and Rivkin (1997).
10. The use of annual earnings, which include money teachers receive for teaching and other occupations, obviously goes beyond comparing pure teaching salaries with salaries in other occupations. I believe that, while more common, using just teaching salaries produces the wrong comparison because teachers enjoy much longer vacations than most other workers. Overall earnings better reflect the monetary benefits of being a teacher as opposed to having a different primary occupation. Broad occupations clearly differ in a variety of nonmonetary ways, including fringe benefits and average length of workday and work year. This analysis assumes that the relative importance of these factors for teaching and other occupations has remained constant over time. Rothstein and Miles (1995) suggest that between 1967 and 1991, benefits for teachers rose faster than those for the rest of the economy, although, as they point out, such comparisons are difficult to make with a high degree of reliability. Private school and public school teachers are grouped together in my analyses; nevertheless, since a roughly constant 10 percent of students attended private school throughout the period, it is unlikely that movement in the earnings of private school teachers would have a significant impact on the overall relative wages of teachers.
11. The clearest evidence comes from a series of covariance, or fixed-effects, estimates of performance differences across teachers (for example, see Hanushek [1971, 1992], Murnane [1975], Murnane and Phillips [1981], and Armor et al. [1976]). These analyses consistently show that the differences between a “good” and a “bad” teacher in the poverty-stricken schools of Gary, Indiana, was approximately one grade level per academic year; that is, a student with a good teacher might advance 1.5 grade equivalents in a school year, while one with a bad teacher might advance only 0.5 grade equivalent (Hanushek 1992). Moreover, the consistency of individual teacher effects across grades and school years indicates that the estimated differences relate directly to teacher quality and not to the specific mix of students and the interaction of teacher and students.

ENDNOTES (*Continued*)

12. The Card and Krueger (1992) study is the most discussed analysis of school resources and earnings, but it follows a larger line of research. For an insightful review of past studies that considers underlying characteristics of the studies, see Betts (1996).

13. Betts (1996) and Hanushek (1997) provide evidence on the generalizability of any conclusions about the effects of resources on measures other than achievement tests. Heckman, Layne-Farrar, and Todd (1996) suggest that the estimates of Card and Krueger are not robust to the database or to the model specification. They also introduce concerns about the validity of assumptions needed to identify the key parameters of the Card and Krueger model. Hanushek, Rivkin, and

Taylor (1996) provide evidence that the aggregated analysis relying on state differences could bias the results toward the finding of resource effects. See, however, the discussion in Card and Krueger (1996).

14. Other factors have entered into the discussion of meta-analysis, but they are not central to the policy discussion here. See Hanushek (1996a, 1996c, 1997).

15. While the decentralization considered here refers to pure resource policies and general funding, the evidence supports this conclusion even at the level of school-based management. See Summers and Johnson (1996).

REFERENCES

- Armor, David, Patricia Conry-Oseguera, Millicent Cox, Niceima King, Lorraine McDonnell, Anthony Pascal, Edward Pauly, and Gail Zellman.* 1976. ANALYSIS OF THE SCHOOL PREFERRED READING PROGRAM IN SELECTED LOS ANGELES MINORITY SCHOOLS. Santa Monica, Calif.: RAND.
- Ballou, Dale, and Michael Podgursky.* 1995. "Recruiting Smarter Teachers." JOURNAL OF HUMAN RESOURCES 30, no. 2 (spring): 326-38.
- Baumol, William J.* 1967. "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis." AMERICAN ECONOMIC REVIEW 57, no. 3 (June): 415-26.
- Berliner, David C., and Bruce J. Biddle.* 1995. THE MANUFACTURED CRISIS: MYTHS, FRAUD, AND THE ATTACK ON AMERICA'S PUBLIC SCHOOLS. Reading, Mass.: Addison-Wesley Publishing Company.
- Betts, Julian R.* 1998. "The Two-Legged Stool: The Neglected Role of Educational Standards in Improving America's Public Schools." Federal Reserve Bank of New York ECONOMIC POLICY REVIEW 4, no. 1.
- . 1996. "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature." In Gary Burtless, ed., DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS, 141-91. Washington, D.C.: Brookings Institution.
- Borland, Melvin V., and Roy M. Howsen.* 1992. "Student Academic Achievement and the Degree of Market Concentration in Education." ECONOMICS OF EDUCATION REVIEW 11, no. 1 (March): 31-9.
- Card, David, and Alan B. Krueger.* 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." JOURNAL OF POLITICAL ECONOMY 100, no. 1 (February): 1-40.
- . 1996. "Labor Market Effects of School Quality: Theory and Evidence." In Gary Burtless, ed., DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS, 97-140. Washington, D.C.: Brookings Institution.
- Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York.* 1966. EQUALITY OF EDUCATIONAL OPPORTUNITY. Washington, D.C.: U.S. Government Printing Office.
- Greenwald, Rob, Larry V. Hedges, and Richard D. Laine.* 1996. "The Effect of School Resources on Student Achievement." REVIEW OF EDUCATIONAL RESEARCH 66, no. 3 (fall): 361-96.
- Hanushek, Eric A.* 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." AMERICAN ECONOMIC REVIEW 60, no. 2 (May): 280-8.
- . 1992. "The Trade-Off between Child Quantity and Quality." JOURNAL OF POLITICAL ECONOMY 100, no. 1 (February): 84-117.
- . 1996a. "A More Complete Picture of School Resource Policies." REVIEW OF EDUCATIONAL RESEARCH 66, no. 3 (fall): 397-409.
- . 1996b. "Rationalizing School Spending: Efficiency, Externalities, and Equity, and Their Connection to Rising Costs." In Victor R. Fuchs, ed., INDIVIDUAL AND SOCIAL RESPONSIBILITY: CHILD CARE, EDUCATION, MEDICAL CARE, AND LONG-TERM CARE IN AMERICA, 59-91. Chicago: University of Chicago Press/NBER.
- . 1996c. "School Resources and Student Performance." In Gary Burtless, ed., DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS, 43-73. Washington, D.C.: Brookings Institution.
- . 1997. "Assessing the Effects of School Resources on Student Performance: An Update." EDUCATIONAL EVALUATION AND POLICY ANALYSIS 19, no. 2 (summer): 141-64.
- Hanushek, Eric A., and Dongwook Kim.* 1996. "Schooling, Labor Force Quality, and the Growth of Nations." Unpublished paper, University of Rochester Department of Economics, December.
- Hanushek, Eric A., and Steven G. Rivkin.* 1997. "Understanding the Twentieth-Century Growth in U.S. School Spending." JOURNAL OF HUMAN RESOURCES 32, no. 1 (winter): 35-68.
- Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor.* 1996. "Aggregation and the Estimated Effects of School Resources." REVIEW OF ECONOMICS AND STATISTICS 78, no. 4 (November): 611-27.
- Hanushek, Eric A., et al.* 1994. MAKING SCHOOLS WORK: IMPROVING PERFORMANCE AND CONTROLLING COSTS. Washington, D.C.: Brookings Institution.
- Heckman, James S., Anne Layne-Farrar, and Petra Todd.* 1996. "Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship." In Gary Burtless, ed., DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS, 192-289. Washington, D.C.: Brookings Institution.

REFERENCES (*Continued*)

- Hoxby, Caroline Minter.* 1996. "How Teachers' Unions Affect Education Production." *QUARTERLY JOURNAL OF ECONOMICS* 111, no. 3 (August): 671-718.
- Krueger, Alan B.* 1998. "Reassessing the View That American Schools Are Broken." Federal Reserve Bank of New York *ECONOMIC POLICY REVIEW* 4, no. 1.
- Murnane, Richard J.* 1975. *IMPACT OF SCHOOL RESOURCES ON THE LEARNING OF INNER CITY CHILDREN.* Cambridge, Mass.: Ballinger.
- Murnane, Richard J., and Barbara Phillips.* 1981. "What Do Effective Teachers of Inner-City Children Have in Common?" *SOCIAL SCIENCE RESEARCH* 10, no. 1 (March): 83-100.
- Peltzman, Sam.* 1993. "The Political Economy of the Decline of American Public Education." *JOURNAL OF LAW AND ECONOMICS* 36 (April): 331-70.
- Poterba, James M.* 1996. "Government Intervention in the Markets for Education and Health Care: How and Why?" In Victor R. Fuchs, ed., *INDIVIDUAL AND SOCIAL RESPONSIBILITY: CHILD CARE, EDUCATION, MEDICAL CARE, AND LONG-TERM CARE IN AMERICA*, 277-304. Chicago: University of Chicago Press/NBER.
- Rothstein, Richard, and Karen Hawley Miles.* 1995. *WHERE'S THE MONEY GONE? CHANGES IN THE LEVEL AND COMPOSITION OF EDUCATION SPENDING.* Washington, D.C.: Economic Policy Institute.
- Summers, Anita A., and Amy W. Johnson.* 1996. "The Effects of School-Based Management Plans." In Eric A. Hanushek and Dale W. Jorgenson, eds., *IMPROVING AMERICA'S SCHOOLS: THE ROLE OF INCENTIVES*, 75-96. Washington, D.C.: National Academy Press.
- U.S. Department of Education.* 1996a. *DIGEST OF EDUCATION STATISTICS 1996.* Washington, D.C.: National Center for Education Statistics.
- . 1996b. *EDUCATION INDICATORS: AN INTERNATIONAL PERSPECTIVE.* Washington, D.C.: National Center for Education Statistics.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

Reassessing the View That American Schools Are Broken

Alan B. Krueger

A growing number of scholars and political commentators have concluded that the U.S. public school system is flawed, and that it can only be corrected by fundamental changes in the institutions that govern education. Chubb and Moe (1990, p. 3), for example, argue that the “existing [educational] institutions cannot solve the problem, because they are the problem.” Widespread belief that the current educational system is flawed, rather than any concrete or systematic evidence indicating that an alternative system performs better than the current one, has motivated frequent calls for radical “institutional reforms” of schools.

The view that the U.S. school system has failed, or is “broken,” is commonly supported by three arguments: (1) there has been a steady decline in the performance of American students on standardized tests, (2) American

children perform worse on international comparisons than foreign children, and (3) the existing system fails to convert school resources (such as smaller classes) into school outputs (such as better test performance).¹

This paper reassesses the claim that American schools are broken. The first section examines trends in National Assessment of Educational Progress (NAEP) test scores, and the relationship between average test score performance and school resources on an aggregate level. Although the aggregate data show a surprisingly strong, *positive* relationship between educational spending and student achievement, one should be cautious about drawing any causal inference from such a relationship because of changes in the composition of students over time and changes in the focus of educational spending. More convincing evidence comes from the randomized experiment on class size, which I describe in the subsequent section. Next, I infer the influence of schooling on student performance by considering gains in student achievement by socioeconomic status (SES) during the school year and during the summer months. The paper’s final

Alan B. Krueger is the Bendheim Professor of Economics and Public Affairs at Princeton University and a research associate of the National Bureau of Economic Research.

section summarizes evidence on the increasing economic rewards associated with completing high school.

The main conclusion from this review is that the widely held belief that American schools have failed—that they are performing worse today than they have in the past, that a high school degree is no longer valuable, and that additional resources yield no benefits in the current system—is not supported by the evidence. The evidence suggests that the perceived crisis in education has been greatly exaggerated, if indeed there is a crisis at all. Nonetheless, major changes in U.S. schooling might produce more desirable results. However, it would not be prudent to radically restructure the U.S. education system out of misplaced frustration that the current system has failed miserably or out of an unsupported presumption that progress cannot be made in the context of the existing system. In light of these findings, the conclusion offers incremental proposals to improve our schools.

WHAT DO THE AGGREGATE ACHIEVEMENT TEST DATA TELL US?

AGGREGATE TIME TRENDS

Concern over the deteriorating performance of U.S. students is often based on time-series trends in the Scholastic Aptitude Test (SAT). For example, Chubb and Moe (1990, pp. 7-8) write, “the single most important symbol of the underlying problem came to be the monotonic decline, from the mid-1960s through 1980, in the scores of high school students on the national Scholastic Aptitude Test, or SAT.” The emphasis on the average SAT score is odd because the exam is not designed to measure students’ current levels of acquired skills, but instead their potential to perform well in college. Even more important, the students who take the SAT are a self-selected lot, and the selection has changed dramatically over time. As a wider segment of American students has attended college, the percentage of high school seniors taking the exam has increased. This increase has been particularly strong among students who rank in the bottom half of their high school class (see Berliner and Biddle [1995]). Because the composition of students taking the SAT has changed over time, the College Entrance Examination Board, which publishes the test, has repeatedly warned against

inferring trends in school or student performance from the SAT (see, for example, College Entrance Examination Board [1988]).

To the extent that one can correct for the changing mix of students who take the SAT, there is little cause for alarm. For example, Berliner and Biddle (p. 22) show that between 1976 and 1993, the average SAT score has gone up for every demographic group except whites, and it declined only slightly for whites. The authors (p. 32) also summarize evidence that shows an upward trend in the 1980s in the California Achievement Test (CAT), the Stanford Achievement Test, the Comprehensive Test of Basic Skills, and other commercial tests. There is little support in these data for the claim made by the National Commission on Excellence in Education (1984, p. 8) that “average achievement of high school students on most standardized tests is now lower than 26 years ago when Sputnik was launched.”

Most analysts probably agree that the NAEP exam provides a more meaningful assessment of trends in student performance than the SAT. Like the SAT, the NAEP is conducted by the Educational Testing Service. But unlike the SAT, the exam is administered to a representative sample of students and is intended to assess progress on basic math, reading, and science skills. The NAEP exam has been administered to nine-, thirteen-, and seventeen-year-olds in selected years since 1970. There are a total of nine time trends that can be analyzed with the NAEP data. Chart 1 presents the average NAEP exam scores for each year, after age and subject main effects have been removed.² For most of the subjects and age groups, the NAEP data display a modest upward time trend after an initial dip in the early 1970s. Indeed, the correlation between the average NAEP score and time (that is, the year in which the test was given) is positive for eight of the nine age-by-subject cases, and it is statistically significant at the 10 percent level for seven of the nine cases. The median of these nine linear trends indicates that test scores are rising by .06 standard deviation per decade.³ It is also possible that the unadjusted NAEP data understate the upward trend in student performance because the composition of students has changed over time. In particular, the rising proportion of students who are immigrants and minorities,

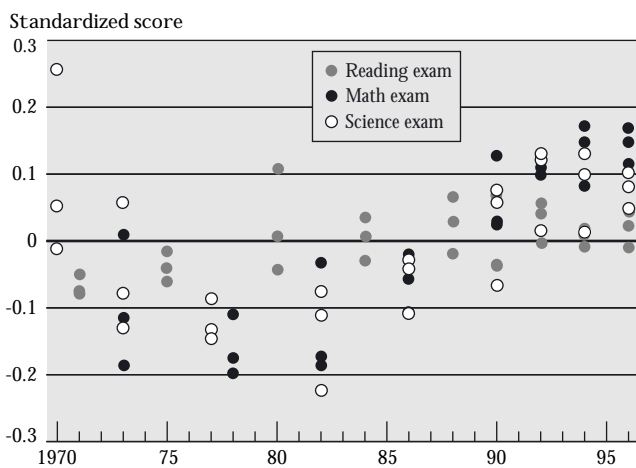
and raised in poverty and by single parents, might be expected to lower average test scores over time.⁴

Chart 2 displays trends in average NAEP mathematics test scores for seventeen-year-old black students and for all students who live in disadvantaged urban communities.⁵ The scores are expressed as deviations from the 1973 overall NAEP score, divided by the 1996 cross-sectional standard deviation. Perhaps surprisingly, the chart shows that the most disadvantaged students have made the greatest gains. The gap in math scores between students in disadvantaged communities and all communities narrowed by approximately one-half of one standard deviation in the 1980s. Moreover, between the early 1970s and 1990, the black-white NAEP mathematics test-score gap for seventeen-year-olds decreased by nearly half, although the gap has expanded in the 1990s. These findings are inconsistent with the popular stereotype that inner-city schools are in decline.

Is the upward trend in the aggregate NAEP scores big or small? To some extent, the significance of the trend is in the eye of the beholder. Hanushek (1996, p. 51), for one, argues that “there is no way to conclude that aggregate performance has increased significantly over the past quarter-century.” The following calculation, however, suggests that

Chart 1

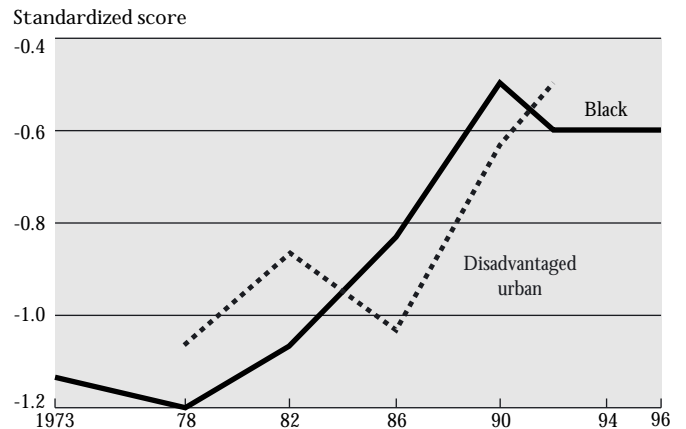
Standardized NAEP Scores over Time



Source: Author's calculations, based on National Center for Education Statistics (1997).
 Note: Each point corresponds to an average adjusted for subject and age group; see endnote 2 for more details.

Chart 2

Standardized NAEP Math Scores for Seventeen-Year-Old Black and Disadvantaged Urban Students



Source: National Center for Education Statistics (1994, 1997).
 Note: Each score was standardized by subtracting the 1973 score for all students and dividing by the 1996 standard deviation across all students.

the time trend is not trivial. Over a twenty-five-year period, the average NAEP score is predicted to have increased by .15 standard deviation, based on the median of the nine linear trends for all subjects and age groups. What does it mean for the average test score to rise by .15 standard deviation? If the distribution of scores is normal, an increase of .15 standard deviation implies that the average (or median) student would have advanced six percentile ranks. In other words, the student scoring in the fiftieth percentile today would perform as well as the fifty-sixth-percentile student did twenty-five years ago. Although this is not a dramatic improvement, it is difficult to find well-evaluated, large-scale educational innovations that have produced equally large gains for the average student.

THE RELATIONSHIP BETWEEN AGGREGATE TEST SCORES AND SCHOOL RESOURCES

Hanushek (1996) presents two notable figures. The first shows a near-exponential growth in expenditures per student from 1890 to 1990. The second shows the average NAEP score for seventeen-year-olds on the math, reading, and science tests for available years since 1970. On the basis of these figures, Hanushek (p. 51) concludes, “the aggregate data provide a *prima-facie* case that school spending and school resources are not linked to performance.”

To my surprise, a straightforward statistical analysis of these data is more supportive of the opposite conclusion (Table 1). Specifically, I pooled the NAEP data across the three age groups and three subject tests and estimated an ordinary least squares (OLS) regression of the form:

$$Y = a + b(\text{spending/student}) + \text{subject dummies} + \text{age dummies},$$

where Y is the average score on the NAEP exam measured in standard deviation units, and spending/student is current school spending in constant 1995-96 dollars divided by the number of enrolled students.⁶ In some specifications, dummy variables are also included for the age of students (nine and thirteen, with seventeen omitted) and for the subject (math and reading, with science omitted).

The regression results in column 1 exclude the subject and age dummies, while the results in column 2 include these explanatory variables. In either model, an increase in expenditures per student has a positive and statistically significant association with average test scores. A \$2,000 increase in expenditures per student is associated with an increase of about .11 standard deviation in the average NAEP score.⁷

The science exam may be more difficult to compare over time than the math or reading exams because of major breakthroughs in basic science in the last thirty years and the diversity of science curricula across schools. In column 5, the sample is limited to the math and reading exams. Here, we find a larger effect of school spending: a \$2,000 increase is associated with an increase of .14 standard deviation in the mean achievement score.

A great deal of work on “educational production functions” has focused on class size. Therefore, in columns 3, 4, and 6 of Table 1 the pupil-teacher ratio is used as a measure of school resources instead of expenditures per student.⁸ These results are also consistent with the view that resources matter. According to the model in column 4, a reduction in the pupil-teacher ratio of eight students—from, say, twenty-three to fifteen—would be associated with an increase in the average score of .176 standard deviation. This is the equivalent of the average student moving up seven percentile ranks, again assuming normality.

To increase the sample size, I pool together all three subject tests and all three age groups in the results reported in Table 1. Perhaps aggregating across age groups and subjects

Table 1
EFFECT OF SCHOOL RESOURCES ON STANDARDIZED NAEP SCORES: POOLED SAMPLE OF MATH, READING, AND SCIENCE SCORES
ACROSS NINE-, THIRTEEN-, AND SEVENTEEN-YEAR-OLDS
OLS Coefficient Estimates with Standard Errors in Parentheses

Variable	(1)	(2)	(3)	(4)	Math and Reading Only	
					(5)	(6)
Spending per enrolled student	.054 (.013)	.057 (.011)	—	—	.072 (.012)	—
Pupil-teacher ratio	—	—	-.015 (.005)	-.022 (.006)	—	-.030 (.007)
Math	—	-.080 (.025)	—	-.078 (.027)	-.132 (.021)	-.131 (.023)
Reading	—	.051 (.024)	—	.051 (.026)	—	—
Age nine	—	-.053 (.024)	—	.032 (.035)	-.060 (.026)	.059 (.038)
Age thirteen	—	-.032 (.024)	—	-.032 (.026)	-.043 (.026)	-.042 (.028)
R ²	.182	.439	.118	.361	.622	.548
Sample size	78	78	78	78	51	51

Sources: National Center for Education Statistics (1997); U.S. Department of Education (1997, Tables 63 and 166).

Notes: Scores have been scaled by subtracting the 1996 score and dividing by the 1996 cross-sectional standard deviation. Spending per enrolled student is in thousands of 1995-96 dollars. Each equation also includes an intercept.

distorts the results. If a separate regression of test scores on expenditures per student is estimated for each of these nine series, however, a positive association is found for eight of the nine cases. Even with a short time series of data, this relationship is statistically significant at the .10 level for seven of the nine cases. The weakest relationships arise for seventeen-year-olds, especially in science (which is negative, with a t-ratio of $-.77$).

I must confess to being surprised by the consistently positive association between NAEP test scores and school resources. To see if my priors were unusual, I e-mailed a short questionnaire to each of the eight other presenters at the “Excellence in Education” conference to assess their expectations of these correlations. Six presenters replied. The median respondent expected six of the nine correlations between NAEP scores and expenditures per student to be positive, with four and a half of the nine correlations statistically significant and positive, and one statistically significant and negative.⁹ Thus, the actual correlations are somewhat more supportive of the view that resources are associated with achievement than this small sample of experts anticipated.

How did Hanushek conclude from the aggregate NAEP data that achievement and school resources are not linked? First, he displayed the NAEP data on a scale ranging from 280 to 310. This is a very wide scale, equivalent to 1 standard deviation on the 1996 NAEP math exam. Under normality, if the average student increased his or her performance by 1 standard deviation, the student would move up thirty-six percentile ranks in the distribution. With such a wide scale, any change in the NAEP score appears visually attenuated. Second, Hanushek only displayed trends for seventeen-year-olds; these students exhibit a weaker relationship between test scores and resources than the other age groups. If the model in column 2 of Table 1 is estimated just for nine- and thirteen-year-olds, for example, the coefficient on spending per student is 33 percent larger.

An analysis of the NAEP scores more thorough than mine—conducted by Grissmer et al. (1997) and based on regional-level data over time—reaches the same qualitative conclusion as that suggested by Table 1. But I

do not wish to extol the findings based on the aggregate NAEP data very much, if at all. Obviously, many relevant factors have changed over time that may bias (either upward or downward) the relationships estimated in Table 1. In addition, to the extent that the generosity of resources is partially determined by low test performance (as in compensatory education), simultaneity bias will attenuate the relationships found in the table. Suffice it to say that my interpretation of the aggregate data is that they provide *prima facie* evidence that student achievement may be linked to school resources. In my view, a far more compelling test of whether resources matter in the current system is discussed in the next section.

THE TENNESSEE STUDENT-TEACHER ACHIEVEMENT RATIO EXPERIMENT

There has been considerable debate over whether devoting more resources to schools in the current system would improve student outcomes.¹⁰ Research has been unable to resolve this debate, in part because it is unclear which variables (family background, innate ability, and so forth) should be held constant when the effect of school resources on student performance is estimated. Additionally, when education production functions are estimated with observational data, there is concern about reverse causality: more resources may be assigned to some schools or classes because of low achievement. Finally, there is no consensus as to the appropriate specification of the education production function. For example, some researchers have related the change in test scores to the level of resources in any given year, some have related the change in test scores to the change in resources, and others have related the level of test scores to the level of resources.

An experiment in which children are randomly assigned to classes with high and low levels of resources would help to overcome many of these statistical problems. Because children are already assigned to teachers and schools, controlled experimentation is more feasible in education than in many other fields. Yet the education field lags behind medicine, job training, and agricultural research in the extent to which controlled experiments are utilized. The Food and Drug Administration requires convincing

evidence from a well-designed experiment before new drugs that influence life and death can be put on the market; but when it comes to new educational innovations, weaker standards of evidence are required.

The Tennessee Student-Teacher Achievement Ratio (STAR) experiment is an exception in the education field. In this experiment, 11,600 Tennessee students in eighty participating schools were randomly assigned to varying sized classes in kindergarten and grades 1 through 3. Mosteller (1995) describes Project STAR as “a controlled experiment which is one of the most important educational investigations ever carried out and illustrates the kind and magnitude of research needed in the field of education to strengthen schools.” Although the experiment was not perfect (what study is?), the results strongly suggest that smaller class sizes help students, especially low-income and minority students.

The key features of the experiment include the following.¹¹ The experiment began in 1986 and included the wave of students who were enrolled in kindergarten that year. Within each participating school, kindergarten students were randomly assigned to a small class (an average of 15.1 students), a regular-size class (an average of 22.4 students), and a regular-size class with a teacher’s aide (an average of 22.8 students). The original plan called for the students to remain in their original class-size assignment until the third grade. After the third grade, the experiment was concluded and all students were assigned to regular-size classes. As noted below, with one important exception, the experiment went largely as planned.

Another feature of the experiment is that additional waves of students entered the experiment in the first grade, the second grade, and the third grade. In particular, because kindergarten attendance was not mandatory in Tennessee at the time of the study, many new students entered the program in the first grade. Moreover, students were added to the sample over time because they repeated a grade or because their families moved to a school zone that included a participating school. Some 2,200 new students entered the project in the first grade and were randomly assigned to the three types of classes. More than 1,000 new students entered the experiment in both the second and the

third grades. Newly entering students were randomly assigned to one of the three class types. This feature of the experiment enables the estimation of class-size effects for each wave of students who entered the experiment in various grades.

The students were given a battery of tests at the end of each school year. I focus on the results of the Stanford Achievement Test. Specifically, I measure student performance by the average percentile rank on the math, reading, and word recognition tests.

The Tennessee STAR experiment is the best designed large-scale educational experiment to date. Nonetheless, it had four important limitations:

- Because of parental complaints, students in the regular-size and regular-size/teacher’s aide classes were randomly reassigned between these two types of classes between kindergarten and first grade, while the students in small classes continued in small classes. Note that results from the kindergarten year are uncontaminated by this deviation from the original experimental design. In addition, my analysis (see Krueger [1997]) suggests that the reassignment of students in regular-size classes in first grade did not invalidate the main results of the experiment.
- The experiment did not collect baseline test scores. These data would have been useful to assess whether the students were uniformly distributed across class types by initial achievement level. Nonetheless, the students’ background characteristics (such as age, race, and probability of receiving free or reduced-price lunch) appear to be uniformly distributed across class types, which suggests that random assignment was carried out successfully.
- In grades 1 through 3, each regular class had the services of a part-time teacher’s aide 25 to 33 percent of the time on average, so the variability in aide services between groups was restricted.¹² Because the present focus of my analysis is primarily on the effect of class size, this feature of the experiment is of less concern.
- Attrition from the sample was high, in part because some students repeated grades and were not tracked, and in part because some students moved to other school districts.¹³

RESULTS OF THE STAR EXPERIMENT

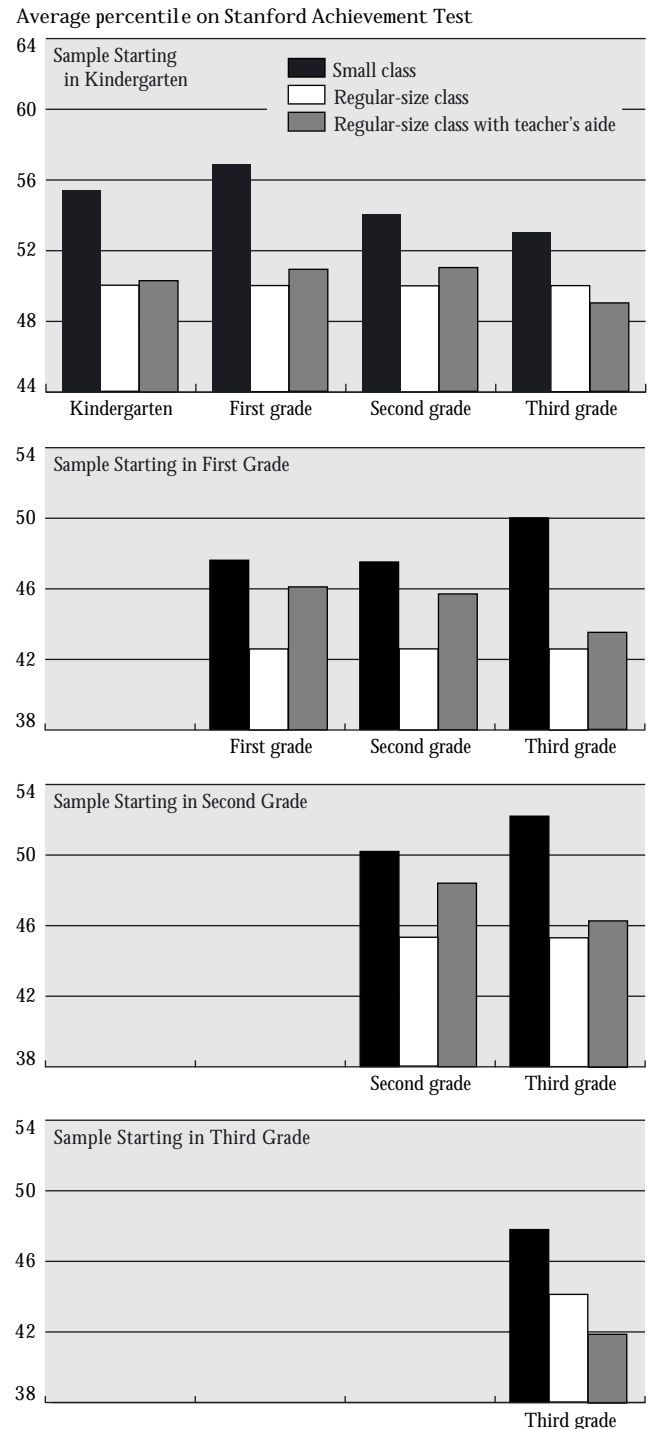
The STAR experiment has been analyzed by Folger and Breda (1989), Word, Johnston, Bain, et al. (1990), Finn and Achilles (1990), and Krueger (1997). The main results of the experiment are summarized below. This summary draws heavily on Krueger, to which the reader is referred for further elaboration of the statistical results.

- The main results for the first four years of the experiment are summarized in Chart 3. For each entry wave of the experiment and grade level, the chart shows the average percentile ranking of students assigned to small classes, regular-size classes, and regular-size classes with a teacher's aide.¹⁴ At the end of the initial year in which students were assigned to small classes, their average performance exceeded that of students in the regular-size and regular-size/teacher's aide classes by about five to eight percentile points. It is an interesting coincidence that this range encompasses the estimated effect of reducing class size by seven to eight students that I obtained from the regression model with the aggregate NAEP data (Table 1).
- For the largest wave, which entered the experiment in kindergarten, the relative advantage of students assigned to small classes tends to grow between kindergarten and first grade, and then is relatively stable. For students who entered in the first or second grade, the advantage of attending a small class tends to grow in subsequent grades.
- In most grades, students assigned to classes with a full-time teacher's aide perform about as well, or only slightly better, than students assigned to regular-size classes without a full-time aide.
- As in all experiments, it is possible that the "treatment" group worked in some way to prove the treatment effective (so-called Hawthorne effects), or that the "control" group worked extra hard to overcome the deficit of being assigned to a small class (so-called John Henry effects).¹⁵ Because there was variability in size even among the classes in the control group, it is possible to explore the likely impact of such "reactive" effects to the experiment. Specifically, I divided the students assigned to regular-size classes into a relatively small class-size group (an average of twenty-one students) and a relatively large class-size group (an average of twenty-five students). I then estimated the difference in average test scores between students in the smaller and larger classes, the results of which appear in Chart 4.¹⁶

Students in the smaller classes among the controls scored higher on the tests than students in the larger classes. Because the benefit of reducing class size is of roughly comparable magnitude in Chart 4 and Chart 3

Chart 3

Effect of Class Size on Test Scores: Tennessee Student-Teacher Achievement Ratio Experiment



Source: Krueger (1997, Figure 2).

and because students (and their teachers) in the smaller classes in Chart 4 did not know they were in a treatment group, there is little support for the view that the main experimental results described earlier are contaminated by Hawthorne effects.

- The effect of attending a smaller class tends to vary systematically across certain groups of students. For example, in the STAR experiment minority students and students on free or reduced-price lunch tended to receive a larger benefit from being assigned to a small class. This pattern is consistent with Summers and Wolfe's (1977) finding that attending a small class is more beneficial for low-achieving students than for high-achieving ones.
- The effect of attending a small class also tends to vary across schools. Notice that in the STAR experiment a separate class-size effect can be estimated for each school, because each school had at least one of each class type. I estimated the effect of attending a small class for each school. The standard deviation of these eighty effects (after adjusting for sampling variability) was 6.9 percentile points. At the average school, the assignment of students to a small class raised performance by 4.6 percentile points. For two-thirds of the schools, the effect of attending a small class was positive, while for one-third it was negative. Furthermore, 30 percent of the schools

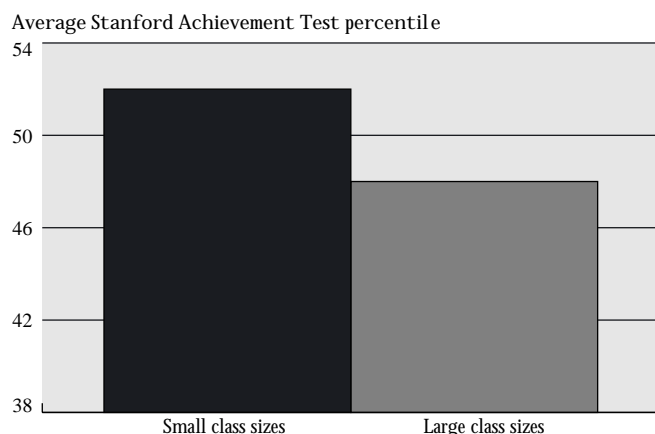
had t-ratios on the small-class effect exceeding 2, while 2.5 percent had t-ratios of less than -2. Smaller classes seem to help student performance at the average school and, indeed, at most schools, although there appears to be a wide distribution of the effect of class size on performance across schools.

If researchers and administrators could determine which schools manage to translate resources more effectively into student performance than others, we could target resources to those types of schools, and try to emulate their practices elsewhere. Consequently, I related the school-level class-size effects to variables such as the racial composition of the students, the urbanicity, and the percentage of students receiving free lunch. Although some of these variables were related to the effect sizes in bivariate regressions, they were all individually insignificant when they were included in a multiple regression.

- The students who participated in Project STAR were returned to regular classes after the third grade and have been tracked since then. Nye, Zaharias, Fulton, et al. (1994) find that students who were placed in small classes had lasting achievement gains through at least the seventh grade, although the later benefits are difficult to compare in magnitude with those at earlier grades because of changes in the tests that were administered. Since the STAR students are currently finishing high school, it would be desirable to learn more about their long-term academic—and just as important, nonacademic—outcomes as they enter early adulthood.

Chart 4

Hawthorne or John Henry Effects: A Comparison of Test Scores within Regular Classes



Source: Author's calculations, based on Tennessee Student-Teacher Achievement Ratio data.

Notes: All grade levels have been pooled together. Small classes have an average of twenty-one students; large classes have an average of twenty-five students.

SUMMER AND SCHOOL-YEAR TEST SCORE GAINS

Another way to infer the impact of schooling on educational achievement is to compare student progress during the school year and during the summer months. Entwisle, Alexander, and Olson (1997) provide a particularly careful application of this approach. Specifically, they collected data on 790 first-time first-grade students from a stratified sample of twenty Baltimore public schools in 1982. These students were tracked for several years. They were given the California Achievement Test at the beginning and end of each school year.¹⁷ Consequently, test score gains could be tracked during the school year and during the summer months when schools were not in session.

Entwisle, Alexander, and Olson divide students into three groups based on their families' socioeconomic status. This categorization involved weighting five standardized variables: mother's and father's occupations and education, and subsidized meal status. The top panel of Chart 5 summarizes their findings concerning seasonal gains on the math CAT for children from the lowest and highest SES groups. The test-score gains during the school year are remarkably similar for the two groups. The bottom panel of Chart 5, however, indicates that during the summer months, children from low-SES families lose ground, while children from high-SES families gain ground. Table 2 summarizes the cumulative gains on the reading and math CATs over the first five years during which the children were tracked. These results suggest that children from high- and low-SES households make roughly the same progress during the school year, although the gap between high- and low-SES children expands during the summer months because

Table 2
CUMULATIVE FIVE-YEAR GAINS ON CAT EXAMS,
BY FAMILY SOCIOECONOMIC STATUS

Socioeconomic Status	Over the School Year	Over the Summer
	Math	
Low	185.8	-8.0
High	186.3	24.9
Reading		
Low	193.3	0.8
High	190.9	46.6

Source: Entwisle, Alexander, and Olson (1997, Table 3.1).

children from low-SES families fall behind when school is not in session.¹⁸ One interpretation of these results is that the public schools overcome whatever learning deficits are associated with low socioeconomic status when schools are in session, but during the summer months the family environmental effects dominate.

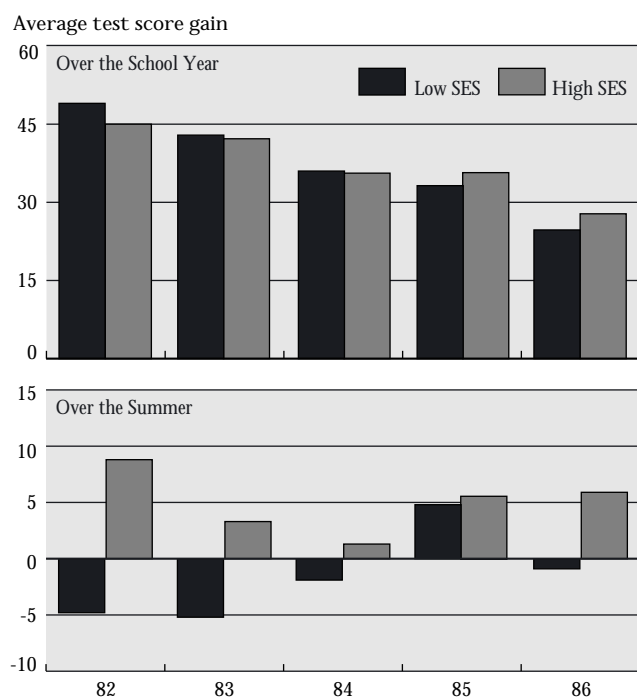
In 1967, Martin Luther King, Jr., wrote, "the job of the school is to teach so well that family background is no longer an issue." At least insofar as test-score gains are concerned, the schools may have achieved this goal. During the school year, students make roughly the same gains regardless of family background. The challenge is to overcome the relative decline in performance of low-income children that occurs when school is not in session, and before school begins.

FINISHING HIGH SCHOOL PAYS OFF

Looking across workers, researchers have found that average earnings tend to be higher for those with more years of schooling.¹⁹ Moreover, a large literature documents that the "monetary return to education"—that is, the earnings of highly educated workers relative to those of less educated workers—rose substantially in the 1980s (see Levy and Murnane [1992] for a survey). Even when the same cohorts are tracked through time, the return to education appears to have increased, suggesting that changes in the skill composition of workers across education levels do not explain this phenomenon. The most compelling interpretation of the rise in the return to education is that the labor market now values skills more highly than it used to, probably because of the combined effects of skill-biased technological changes, globalization, and institutional changes.

Chart 5

Seasonal Gains on CAT Math Exam, by Family Socioeconomic Status

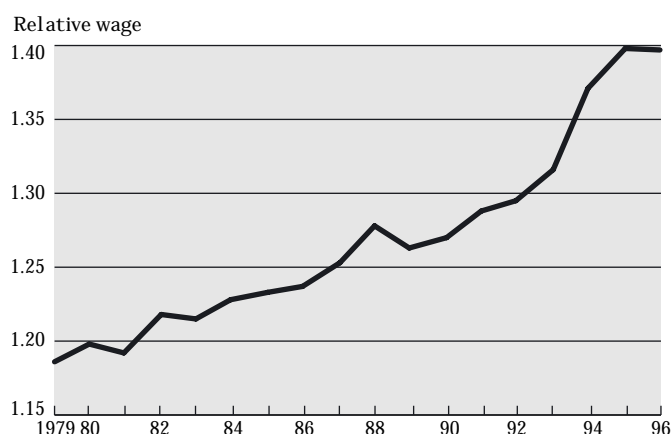


Source: Entwisle, Alexander, and Olson (1997, Table 3.1).

Notes: Sample consists of 498 Baltimore public school students who entered first grade in 1982. SES = socioeconomic status.

Chart 6

Ratio of Median Weekly Wage of High School Graduates to that of High School Dropouts



Source: Author's calculations, based on unpublished Bureau of Labor Statistics tables.

Note: The chart presents the ratio of median weekly wages of workers with exactly a high school education to those with less than a high school education for full-time workers aged twenty-five and older.

Although much of the literature on the return to education focuses on the college–high school wage ratio, Chart 6 documents that it is also true that the ratio of earnings of those with exactly a high school degree to those with less than a high school degree increased over time. Between 1979 and 1996, for example, the earnings advantage of high school graduates relative to that of high school dropouts roughly doubled, from 19 percent to 40 percent. Although the level of earnings of high school graduates has declined over the last two decades once we adjust for inflation, the relative earnings of high school graduates has increased substantially. This finding suggests that—compared with the alternative—completing high school is more valuable today than it used to be, which is inconsistent with the view that the secondary education system fails to produce something of value.²⁰

CONCLUSION

The evidence reviewed in this paper suggests that the U.S. public school system has not deteriorated and may in fact be reasonably efficacious. It is nonetheless possible that school resources could be deployed more effectively and educational innovations could improve student outcomes. Moreover, the seemingly ever-increasing demand for skilled workers in the

United States underscores the potential economic benefits of enhancing the skills of future generations of Americans by improving public schooling. But the conclusion that the current educational system is not inherently flawed leads to policy proposals of a more incremental nature. My personal view is that policymakers should be risk-averse when it comes to changing the public school system. To alter the institutional structure of U.S. schools radically without sufficient evidence that the “reforms” would be successful is to put our children at risk. With this in mind, I would propose that the following incremental reforms be given serious consideration.

- Careful experimentation and evaluation should proceed on a limited basis before wide-scale institutional changes are introduced. To the extent possible, randomized assignment to treatment and control groups should be the research design used to evaluate educational initiatives, such as vouchers, magnet schools, and charter schools. Long-term follow-up focusing on concrete outcomes, not just test scores, would also be desirable. The research base on which educational policy is formulated should be greatly expanded. More experimentation and evaluation are needed. A risk-averse strategy would require fairly convincing evidence that any new educational initiative is effective before it is implemented on a wide scale.
- Demographic projections indicate that public school enrollment will rise in the coming decade. This rise, in turn, will lead to pressure to increase class size or school spending. The research suggests to me that an increase in class size, especially in the early grades, would lower the average student's performance. Society may choose to increase class sizes rather than incur additional educational expenses, but this decision should be made with the expectation that student achievement will be affected by the level of school resources. Decision makers and the public should not bank on there being a “free lunch” when it comes to raising class size.
- In most U.S. public schools, the school year runs about 180 days. This is shorter than in many other industrialized countries. One of the strongest findings in the economics of education literature is that, on average, students who attend school longer earn higher incomes in the labor market. It is reasonable to hypothesize that more days of schooling would have a similar effect. In addition, Entwisle, Alexander, and

Olson's (1997) work suggests that disadvantaged children fall behind in the summer months when schools are closed. I would propose experimenting with an expanded school year of, say, 210 days. To facilitate evaluation of such a change, it would be desirable to phase in an extension of the school year across randomly selected communities within a state.

- The school day in the United States is a legacy of a bygone era when most women did not work for pay.

Times have changed. Probably because of lax supervision in the afternoon when most parents are working, the juvenile violent crime rate peaks between 2 p.m. and 4 p.m. on school days.²¹ On nonschool days, the crime rate peaks in the evening. This finding suggests possible gains from extending the school day, or from providing a richer set of after-school activities. Again, experimentation and evaluation would be desirable.

ENDNOTES

The author thanks Michele McLaughlin for very helpful research assistance and Lisa Barrow, Michele McLaughlin, and Cecilia Rouse for helpful comments.

1. For example, these arguments are expressed in National Commission on Excellence in Education (1984) and Chubb and Moe (1990).
2. The NAEP scores in the chart were first standardized by subtracting the 1996 test score (for that age group and subject) from each observation, then dividing this quantity by the 1996 cross-sectional standard deviation. The transformed scores were then regressed on two age dummies and two subject dummies. The residuals from this regression are displayed. This regression adjustment was used to remove average age and subject effects.
3. The linear trends were estimated from a bivariate regression of the NAEP exam on the year in which the test was taken. The NAEP score is scaled relative to the cross-sectional standard deviation on the exam in 1996. See also National Center for Education Statistics (1997) for a discussion of trends in NAEP scores.
4. However, these factors may be partially or fully offset by the rise in average educational attainment of parents and the decline in the number of children per family.
5. Disadvantaged urban students are defined as those who live in metropolitan statistical areas and attend schools where a high proportion of the students' parents are on welfare or are not regularly employed. Unfortunately, NAEP data are available for this group only from 1978 to 1992.
6. More precisely, if the average test score in year t is Y_t , the average score in 1996 for that subject and age group is Y_{96} , and the standard deviation of scores across students in 1996 for that subject and age group is s_{96} , then the dependent variable is $(Y_t - Y_{96})/s_{96}$. Thus, each year's test score is measured in standard deviation units from the 1996 average test score for that age group and subject.
7. Hanushek (1996) focuses on expenditures per pupil in average daily attendance instead of expenditures per enrolled pupil. Because enrollment is arguably less endogenously determined than attendance, I use expenditures per enrolled student. If the attendance measure is used, however, the coefficient in Table 1, column 2, row 1 is similar: .053 (standard error of .010).
8. A number of arguments could be made that the pupil-teacher ratio is a superior measure for these specifications. Expenditures will be heavily influenced by teacher pay, which varies over time in part because of external labor market forces that influence the pay of college-trained women. Perhaps more important, in these results the pupil-teacher ratio is more closely tailored to the grade level of the students, whereas the expenditures per student pool all grade levels. But I would not want to identify any particular estimate in Table 1 as "the" correct estimate.
9. Another question on my survey was, "Suppose that public school spending in the U.S. were to permanently increase by \$2,000 per student in 1997 dollars. Give your best point estimate of the probability that student achievement would increase, on average." The median respondent thought there was a 75 percent probability that average achievement would rise in this hypothetical situation. Some of the respondents indicated that they expected any increase to be small, however.
10. Even quantitative summaries of the literature reach contrasting findings (see, for example, Hanushek [1986] and Hedges, Laine, and Greenwald [1994]).
11. Project STAR was funded by the Tennessee legislature, at a total cost of approximately \$12 million. The research was designed and carried out by research teams from Tennessee State University, Memphis State University, the University of Tennessee, and Vanderbilt University. See Word, Johnston, Bain, et al. (1990) and Folger (1989) for detailed descriptions of the experiment.
12. Regular classes often had a teacher's aide because the ethic underlying the study was that students in the control group (the regular classes) would not be prevented from receiving resources that they would ordinarily receive.
13. Straightforward adjustments for attrition (for instance, assigning the last test score to students who leave the sample for each subsequent year) did not indicate that attrition distorted the results (Krueger 1997).
14. The results are based on separate OLS regression models estimated for each entry wave and grade level. The dependent variable in the regression model is the student's percentile ranking on the Stanford Achievement Test, and the independent variables are class-type assignment dummies, school dummy variables, race, sex, free-lunch status, teacher sex, teacher race, and teacher education (see Krueger [1997, Table 6, column 4]).
15. In the current case, one may be more concerned about Hawthorne effects because the treatment appears to have been effective, although John Henry effects may lead to an underestimate of the benefit of attending a small class. For an interesting study that casts doubt on the presence of Hawthorne effects in the original Hawthorne experiments, see Jones (1992).

ENDNOTES (*Continued*)

16. To increase the sample size, I pooled students across all grade levels and entry waves. The results shown in Chart 4 are from a regression of the students' percentile rank on a dummy variable for class size, student characteristics, and teacher characteristics.

17. Students were tested regardless of whether they were held back a grade. In principle, the CAT is scaled so that scores are comparable across grade levels.

18. Murnane (1975) similarly finds that test scores for inner-city children increase during the school year and either stagnate or decline during the summer months. Grossman and Sipe (1992) find that achievement levels remain constant for disadvantaged youth who are randomly selected to participate in summer school, while test scores decline for the control group over the summer. They find no long-term benefits of the summer program, however.

19. A long-standing concern has been that the observed earnings-education gradient might reflect unobserved factors such as inherent ability or family background. Most of the literature that tries to estimate the payoff to years of education, however, concludes that omitted variables do not seriously bias the OLS estimates (see, for example, Griliches [1977] and Angrist and Krueger [1991]).

20. For studies of how school resources relate to the return to education, see Card and Krueger (1992) and Heckman, Layne-Farrar, and Todd (1996).

21. See, for example, U.S. Department of Justice (1996), which is based on FBI data on violent crime in South Carolina.

REFERENCES

- Angrist, Joshua, and Alan Krueger.* 1991. "Does Compulsory Schooling Affect Schooling and Earnings?" *QUARTERLY JOURNAL OF ECONOMICS* 106 (November): 979-1014.
- Berliner, David C., and Bruce J. Biddle.* 1995. *THE MANUFACTURED CRISIS: MYTHS, FRAUD, AND THE ATTACK ON AMERICA'S PUBLIC SCHOOLS.* Reading, Mass.: Addison-Wesley Publishing Company.
- Card, David, and Alan B. Krueger.* 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *JOURNAL OF POLITICAL ECONOMY* 100 (February): 1-40.
- Chubb, John E., and Terry M. Moe.* 1990. *POLITICS, MARKETS AND AMERICA'S SCHOOLS.* Washington, D.C.: Brookings Institution.
- College Entrance Examination Board.* 1988. *GUIDELINES ON THE USES OF COLLEGE BOARD TEST SCORES AND RELATED DATA.* New York.
- Entwisle, Doris, Karl Alexander, and Linda Olson.* 1997. *CHILDREN, SCHOOLS AND INEQUALITY.* Boulder, Colo.: Westview Press.
- Finn, Jeremy D., and Charles M. Achilles.* 1990. "Answers and Questions About Class Size: A Statewide Experiment." *AMERICAN EDUCATIONAL RESEARCH JOURNAL* 27 (fall): 557-77.
- Folger, John.* 1989. "Editor's Introduction: Project STAR and Class Size Policy." *PEABODY JOURNAL OF EDUCATION* 67, no. 1 (fall): 1-16.
- Folger, John, and Carolyn Breda.* 1989. "Evidence From Project STAR About Class Size and Student Achievement." *PEABODY JOURNAL OF EDUCATION* 67, no. 1 (fall): 17-33.
- Griliches, Zvi.* 1977. "Estimating the Returns to Schooling—Some Econometric Problems." *ECONOMETRICA* 45 (January): 1-22.
- Grissmer, David, Stephanie Williamson, Sheila Kirby, and Mark Berends.* 1997. "Exploring the Rapid Rise in Black Achievement Scores in the United States (1970-1990)." Unpublished paper, RAND Institute on Education and Training, Washington, D.C.
- Grossman, Jean Baldwin, and Cynthia L. Sipe.* 1992. "Summer Training and Education Program (STEP): Report on Long-term Impacts." Philadelphia: Public Private Ventures, winter.
- Hanushek, Eric A.* 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *JOURNAL OF ECONOMIC LITERATURE* 24 (September): 1141-77.
- . 1996. "School Resources and Student Performance." In Gary Burtless, ed., *DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS*, 43-73. Washington, D.C.: Brookings Institution.
- Hekman, James J., Anne Layne-Farrar, and Petra Todd.* 1996. "Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship." In Gary Burtless, ed., *THE LINK BETWEEN SCHOOLS, STUDENT ACHIEVEMENT, AND ADULT SUCCESS.* Washington, D.C.: Brookings Institution.
- Hedges, Larry V., Richard Laine, and Rob Greenwald.* 1994. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *EDUCATION RESEARCHER* 23, no. 3 (April): 5-14.
- Jones, Stephen.* 1992. "Was There a Hawthorne Effect?" *AMERICAN JOURNAL OF SOCIOLOGY* 98, no. 3 (November): 451-68.
- Krueger, Alan B.* 1997. "Experimental Estimates of Educational Production Functions." NBER Working Paper no. 6051.
- Levy, Frank, and Richard Murnane.* 1992. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations." *JOURNAL OF ECONOMIC LITERATURE* 30, no. 3 (September): 1333-81.
- Mosteller, Frederick.* 1995. "The Tennessee Study of Class Size in the Early School Grades." *THE FUTURE OF CHILDREN: CRITICAL ISSUES FOR CHILDREN AND YOUTHS* 5 (summer/fall): 113-27.
- Murnane, Richard.* 1975. *THE IMPACT OF SCHOOL RESOURCES ON THE LEARNING OF INNER CITY CHILDREN.* Cambridge, Mass.: Ballinger Publishing Company.
- National Center for Education Statistics.* 1994. *NAEP 1992 TRENDS IN ACADEMIC PROGRESS.* Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, July.
- . 1997. *NAEP 1996 TRENDS IN ACADEMIC PROGRESS.* Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, August.
- National Commission on Excellence in Education.* 1984. *A NATION AT RISK: THE IMPERATIVES FOR EDUCATIONAL REFORM.* Washington, D.C.: U.S. Department of Education.

REFERENCES (*Continued*)

Nye, Barbara, Jayne Zaharias, B.D. Fulton, et al. 1994. THE LASTING BENEFITS STUDY: A CONTINUING ANALYSIS OF THE EFFECT OF SMALL CLASS SIZE IN KINDERGARTEN THROUGH THIRD GRADE ON STUDENT ACHIEVEMENT TEST SCORES IN SUBSEQUENT GRADE LEVELS. Seventh-grade technical report. Nashville, Tenn.: Center of Excellence for Research in Basic Skills, Tennessee State University.

Summers, Anita A., and Barbara L. Wolfe. 1977. "Do Schools Make a Difference?" *AMERICAN ECONOMIC REVIEW* 67, no. 4 (September): 639-52.

U.S. Department of Education. 1997. DIGEST OF EDUCATION STATISTICS 1996. Washington, D.C.

U.S. Department of Justice. 1996. JUVENILE OFFENDERS AND VICTIMS: 1996 UPDATE ON VIOLENCE. Washington, D.C.: Office of Juvenile Justice and Delinquency Prevention.

Word, Elizabeth, J. Johnston, Helen Bain, et al. 1990. "The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report, 1985-1990." Nashville, Tenn.: Tennessee State Department of Education.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

Summary of Floor Discussion

Carol Rapaport

The discussion following the session on educational resources and outcomes focused on three issues: the measurement of educational inputs and outputs, changes in the educational incentive structures, and the role of the classroom environment.

Frank Levy introduced the topic of educational input and output measurement by observing that maximizing student achievement has not always been teachers' primary objective. For example, teachers have also sought to provide drug education and to prevent students from dropping out. Chris Meyer proposed using housing prices as an alternative measure of educational output. According to Meyer's research on the effects of tax limitation legislation in Massachusetts, parents do value educational expenditures, because a district's ability to spend money is factored into housing prices. Introducing a related point, Derek Neal asked how much the cost of hiring and employing teachers has changed over the past thirty years. The consensus was that less than half the rise in per pupil expenditures since 1970 has come from a rise in teacher costs.

Participants then turned to the subject of changes in educational incentive structures. Raquel Fernandez wondered why students' and schools' incentives to achieve are weaker now than in the past; she also asked why the educational incentive structure is weaker in the United States than in many other countries. Alan Krueger expressed a contrary view, asserting that the incentives for

students to achieve are in fact stronger than in the past. Julian Betts noted that the real wages of high school dropouts have plummeted. Betts' general outlook was positive—the incentive structure within schools is moving in the right direction. Krueger also argued that schools themselves are given incentives to produce positive student outputs such as good citizenship and preparation for work.

Caroline Hoxby observed that incentive issues complicated the interpretation of an experiment in Tennessee with class size. Studies of the Tennessee Student-Teacher Achievement Ratio (STAR) experiment have concluded that reductions in class size do increase achievement. The Tennessee teachers, however, knew that they were being observed and that the experimental outcomes would inform future policy. Arguably, the teachers in the experiment thus had an incentive to improve student achievement. Krueger, however, pointed out that each individual teacher still had the opportunity to “free ride.” In other words, even if one teacher worked harder, overall achievement averages would not change. The nonexperimental Tennessee data also show that class size affects achievement.

Turning to international comparisons of incentives, Eric Hanushek emphasized that in some countries the goal of students in the primary and secondary grades is to gain admission to a university. For many of these countries, it is difficult to disentangle the effects of public school education, private school education, and family background on student outcomes. Nonetheless,

Krueger believes that a relationship between resources and student achievement exists across countries. Betts stressed the importance of high school graduation exams as a motivating factor for students outside the United States.

In the final part of the discussion, Bill Andrews asked whether the speakers had considered classroom environment. Hanushek stated that the classroom environment mattered to achievement because of the large skill differences among teachers. Although showering money on schools might cause some change in the classroom environment, he argued that it would not bring about a big improvement in student performance. Krueger reminded

participants that the experimental Tennessee data show that class size affects student performance. In addition, peer groups matter: high-achieving students benefit from high-achieving classmates. For weaker students, the quality of classmates did not affect performance; for these same students, however, smaller classes did influence performance.

Krueger concluded the discussion, stating that both class size and expenditures affect student performance. Acknowledging that not all researchers have come to this conclusion, he suggested that many of the studies with results to the contrary suffer from misspecified models.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

What Do America's "Traditional" Forms of School Choice Teach Us about School Choice Reforms?

Caroline M. Hoxby

The majority of U.S. states are currently considering or have recently passed reforms that increase the ease with which parents can choose a school for their children (Tucker and Lauber 1995). At first view, these reforms seem to take elementary and secondary education into wholly unknown territory. Yet this view neglects the fact that choices made by American parents have traditionally been an important force in determining the education their children receive. Parents' ability to choose among fiscally independent public school districts (through residential decisions) and to choose private schools (by paying tuition) is such an established feature of American education that it is almost taken for granted. Yet, through these choices, American parents exercise more control over their children's schooling than do many of their European counterparts. Of

course, American parents are not all equally able to exercise choice. High-income parents routinely exercise more choice because they have more school districts and private schools within their choice sets. In addition, there is significant variation in the degree of choice across different areas of the country. Some metropolitan areas, for instance, contain many independent school districts and/or a number of private schools. Other metropolitan areas are completely monopolized by one school district or have almost no private schooling.

The purpose of this paper is to answer three related questions. First, what general facts can we learn by examining the traditional forms of school choice in the United States? In particular, we need to understand the general relationship between school choice and five factors: (1) student achievement, (2) student segregation (along lines of ability, income, and taste for education, as well as race and ethnicity),¹ (3) school efficiency, (4) teachers' salaries and teacher unionism, and (5) the degree to which parents are involved in and influence their children's

Caroline M. Hoxby is an associate professor of economics at Harvard University and a faculty research fellow of the National Bureau of Economic Research.

schools. Second, how do the general facts that we garner from traditional school choice carry over to analyses of reforms such as charter schools, vouchers for private schools, and open enrollment programs? Third, what information do we still need if we are to predict accurately the effects of reforms? And, what empirical strategies might we pursue to get such information?

For evidence, I draw upon previous empirical work contained in several studies.² Although I briefly sketch the empirical strategy of these studies, this paper does not attempt to present the results in detail. Rather, the goal is to summarize the results and discuss their implications for school choice reforms.

TRADITIONAL CHOICE AND THE ISSUES

There are two basic forms of school choice in the United States. The first is choice among public school districts that have a substantial degree of fiscal and administrative autonomy. The second is choice between public and private schools. In this section, I take each in turn. Later, I briefly discuss intradistrict choice—a scheme that contains some characteristics of the two basic forms of choice.

TRADITIONAL CHOICE AMONG PUBLIC SCHOOL DISTRICTS

Households choose among public school districts by selecting a residence. The degree to which households can exercise this form of choice depends heavily on the number, size, and residence patterns of the school districts in the area centered around their jobs. Some metropolitan areas in the United States have many small school districts with reasonably comparable characteristics. Boston, for instance, has seventy school districts within a thirty-minute commute of the downtown area and many more within a forty-five-minute commute. Miami, on the other hand, has only one school district (Dade County) that covers the entire metropolitan area. People with jobs in rural areas typically have only one or a few alternative school districts to choose from.

This form of choice (among public school districts) has several important properties. First, districts that are good, efficient providers of schooling tend to be

rewarded with larger budgets. This fiscal reward process works because a district's budget nearly always depends on property taxes, which in turn depend on home prices within the district, which in turn depend on how the marginal home buyer values the local schools. Rewards for good, efficient provision occur as long as districts have a significant amount of fiscal autonomy (especially over marginal revenues and expenditures).³ The fiscal reward process tends to be sustainable over the long term because it depends on decentralized choices. This process is in contrast to centralized reward systems—for example, financial or other “merit” awards for successful school districts that are distributed by the state. These systems tend to be unsustainable because states cannot, *ex post*, credibly adhere to processes that reduce (in relative terms) the amount of money going to failing school districts.

The second important property of traditional choice among public school districts is that parents who prefer different amounts of school spending and different types of schools sort themselves into different districts. As a result, each district tends to be more homogeneous than the area, and the residents of each district tend to vote for taxes and schools that approximately fulfill their spending and curriculum desires. This means that districts offer somewhat differentiated schooling that follows local parents' preferences to a certain degree.

Consequently, choice among public school districts creates residential patterns (residential segregation) that mirror households' desired levels of school spending. This result is in contrast to residential patterns that purely reflect households' incomes or housing desires. Of course, desired school spending depends partly on income, but it also depends on the extent to which a household prefers to spend money on schooling relative to other goods or investments. Low-income and minority households are the most likely to be prevented from making reasonably optimal investments in their children's schooling. The ability of these households to choose residences in more than one district may be severely constrained by their budgets or by discrimination.

Another consequence of choice among public school districts is that parents' preferences have some influ-

ence over local schools. Any given school district budget, for instance, is allocated more according to parents' preferences (than, say, according to the preferences of school staff or the state department of education) when parents have more choice among districts. This is true simply because when parents have more choices, school budgets are more elastic with respect to parents' preferences. Therefore, policy is more responsive to those preferences.

Evidence of what happens when an area has more choice among public school districts is useful mainly for analyzing charter school reforms and open enrollment reforms. A charter school receives a charter to educate public school students and a "tuition" payment (from public revenues) for each pupil it enrolls. The school admits students nonselectively or at random. Charter schools are supposed to have a high degree of administrative autonomy from local public schools and to have as much fiscal autonomy as a stable tuition payment per pupil can give them.⁴ Thus, opening a charter school has some, but not all, of the features of creating additional public school districts to compete with the initial district.

An open enrollment program allows students to attend schools in districts outside their districts of residence. Whether or not an open enrollment program closely resembles an expansion of choice among public school districts depends largely on the financial transfers that accompany transferring students. If a program has financial transfers that closely simulate the fiscal pressures of choice among public school districts, it can be regarded as a means of intensifying traditional choice among public school districts by reducing mobility costs and by allowing many more households to be on the margin between districts. Most open enrollment programs, however, have financial arrangements that do *not* simulate the fiscal pressures of choice among districts. For instance, the transfer is often small compared with the receiving district's average expenditure per pupil. Also, the money that accompanies the transferring student often comes wholly or partly from the state rather than from the sending district.

In summary, traditional choice among public school districts is helpful for analyzing charter school and open enrollment reforms because all three types of choice

give us a general sense of (1) the bases on which parents choose among schools, (2) how public schools differentiate themselves given that they are all subject to public scrutiny and public constraints, (3) whether public providers can and do react to competition for students by improving their programs, (4) how the degree of choice among public providers affects parents' willingness to pay for private school alternatives, and (5) how students self-segregate among schools when they are given more choices at the same time that the receiving schools cannot discriminate among them.⁵ Traditional choice among public school districts is less helpful for understanding charter school and open enrollment reforms to the extent that the financial arrangements of the reforms have quite different properties than traditional choice. In addition, charter schools and open enrollment programs depend on the sufferance or cooperation of local school districts, making them less sustainable than traditional choice programs.

TRADITIONAL CHOICE BETWEEN PUBLIC AND PRIVATE SCHOOLS

The second way in which parents have traditionally been able to exercise choice in the United States is by enrolling their children in private schools. Private school tuition in America is not subsidized by public monies (as it is in some European countries), so parents can afford private school only if they can pay both tuition and local taxes supporting public schools.⁶ Partly as a result, private schools tend to enroll fewer than 15 percent of American elementary and secondary students. This percentage reached a peak of just under 15 percent in the early 1960s. Although it declined to 10 percent by 1980, it has since rebounded to 12 percent.

There is tremendous variation in the schooling offered in and tuition paid for private schools in the United States. Approximately 90 percent of private school students attend a school that is affiliated with a religious group, including a variety of Christian and non-Christian groups. Tuition for these schools ranges from a token amount ("\$100 or whatever parents can pay") to more than \$10,000. The remaining 10 percent of private school students attend schools with no religious affiliation; these

institutions include many of the college-preparatory, “independent” schools that charge tuition of \$5,000 or more. More than 65 percent of U.S. private school students attend a school affiliated with the Catholic Church; these institutions vary from modest parochial schools asking for token tuition to elite, college-preparatory schools that compete with the “independents” for students. The modal private school student in the United States attends a Catholic school that is parochial or diocesan and charges a tuition of about \$800 (for elementary school) or \$2,000 (for secondary school).

A key feature of American private schools is that they typically subsidize tuition with monies from donations or (less often) endowment income. The share of schooling costs that is covered by subsidies is larger in schools that serve low-income students; even relatively expensive private schools, however, charge subsidized tuitions. For instance, Catholic elementary schools, on average, cover 50 percent of their costs with donations from local households and the local diocese (they are also implicitly subsidized by teachers who are members of religious orders and accept minimal payment). Catholic secondary school tuitions are less subsidized: on average, they represent about 75 percent of the actual costs of schooling. Even the most expensive religiously affiliated private schools in the United States—schools affiliated with the Friends (Quakers)—charge tuitions that average only 80 percent of their costs.⁷ Note that schools that serve low-income households and charge highly subsidized tuitions are frequently oversubscribed and must ration school places using waiting lists.

Some cities and areas of the United States have significantly larger shares of students in private schools than do others. The shares for metropolitan areas, for instance, range from 35 percent to roughly 0 percent. This variation is created by historical accident, by the donations available for subsidizing private schools in an area, and by the quality of public schools. I return to these sources of variation below.

Choice between private and public schools has several important properties. First, private schools that efficiently offer high-quality education tend to be rewarded

by gaining more applicants. At the very least, the larger applicant pool allows the private school to be more selective. More often, a larger applicant pool allows a private school to expand. Symmetrically, public schools that do not offer quality education efficiently are likely to lose students to private schools. The students who are drawn away are, for any given public school, those with the greatest taste for the type of education offered by private schools. Second, private schools are likely to have an ambiguous impact on the finances of local public schools. On the one hand, an increased supply of private schools tends to draw into the private school sector many parents who might have supported generous public school spending if their children had remained in public schools. This phenomenon tends to decrease voter support for public school spending. On the other hand, an increased supply of private schools tends to draw into the private sector many students who otherwise would have had to be educated at public expense. This phenomenon tends to increase public school spending *per pupil*.

An increase in private school availability should change patterns of residential segregation for the following reasons: private school parents who would choose to live in districts with expensive public schools if private schools did not exist would be willing to live in less expensive districts. Such changes in residential segregation, however, are limited by the fact that private school parents prefer to live with neighbors who have similar professions, educations, and preferences for other local public goods. For instance, private school parents are unlikely to live with low-income neighbors just to avoid paying taxes to support moderately expensive public schools. Finally, private schools put mild pressure on public schools to pay the same input costs that they (private schools) pay. In particular, private schools are less likely to be unionized and to accept supply contracts for political reasons. If they do not pay union wage premiums and are charged competitive prices for supplies, their lower costs indirectly put a little pressure on public schools to be cost efficient. The pressure is small because the fact that private school parents continue to pay taxes to support public schools drives a considerable price wedge between private and public schools with comparable costs.

Evidence about the effects of traditional private school choice is most useful for predicting the effects of vouchers. Some properties of vouchers are quite similar to those of traditional private school choice: successful private schools are rewarded with larger pools of applicants; the least efficient public schools are the most likely to lose students. The fiscal impact of vouchers on public schools is ambiguous, although it is possibly less positive than the fiscal impact of private school competition on public schools. The difference is that vouchers are typically funded with monies from the local public schools. Some students who would attend private schools even in the absence of a voucher program will use vouchers: this fact will have a negative impact on per pupil spending in the sending district. This effect, however, will be offset by the positive impact on per pupil spending that occurs whenever a voucher is used by a student who would have, in the absence of a voucher program, attended a public school. This positive impact occurs because all voucher amounts proposed thus far have been significantly smaller than per pupil spending in the sending public school district. Some of the indirect fiscal impacts of vouchers on per pupil public school spending are positive as well. For instance, some parents with a high taste for education are likely to remain in districts that they would have abandoned for suburban districts if vouchers were not available. Keeping these parents has a positive effect on a district's property prices and, thus, on the tax base that supports public schools.

INTERACTIONS BETWEEN THE TWO TRADITIONAL FORMS OF SCHOOL CHOICE

We expect that the two traditional forms of school choice will substitute for one another to some degree. Parents who can choose a district that offers schooling and a per pupil cost closer to their desires have less incentive to send their children to private school. Of course, public and private school choice are unlikely to substitute for one another completely because the two sectors function under somewhat different constraints. For instance, parents with strong preferences for religious education cannot satisfy such preferences in the public sector; parents with strong pref-

erences for public schooling cannot satisfy such preferences in private schools.

Similarly, we expect some interaction among the reforms. Availability of charter schools is likely to reduce the use of private school vouchers or open enrollment programs. Logically, the more one reform offers a needed type of choice, the less the alternative reforms will be desired or used. For instance, the less autonomous a charter school is, the more parents will want to use private school vouchers. In addition, areas that already have substantial amounts of choice among public school districts or choice of private schools are unlikely to make heavy use of a charter school program or an open enrollment program (unless the latter has perverse fiscal arrangements). Also, areas with substantial amounts of choice among public school districts are less likely to make heavy use of vouchers. The same cannot be said of areas that already have substantial amounts of private school choice. Since vouchers give a transfer to any parent already using private schools, vouchers would be highly utilized in areas with high shares of private schools. The means testing in most proposed voucher programs attempts to reduce transfers to parents already using private schools.

EVIDENCE ON THE EFFECTS OF COMPETITION AMONG PUBLIC SCHOOL DISTRICTS

To determine the effects of competition among public schools, we might compare metropolitan areas that have had long-term differences in parents' ease of choice among districts.⁸ Ease of choice depends on both the number of districts in the area and the evenness with which enrollment is spread over those districts. Choice is easier in a metropolitan area where parents choose among twenty districts of equal size than in an area where three-quarters of enrollment falls into one of twenty districts; choice in the latter area is easier than in an area with only one school district. The inverse of a Herfindahl index based on districts' enrollment shares is a good measure of the ease of choice because it incorporates both these facts—the number of districts and evenness of districts' enrollment shares.⁹

The sizable differences between metropolitan areas in the amount of choice available are largely a result of historical accident and geography. However, we should consider that districts' enrollments can reflect their success: a highly successful and efficient district might attract a disproportionate share of its metropolitan area's enrollment. It might even attract smaller districts to consolidate with it. These phenomena would tend to make simple comparisons of metropolitan areas with public school enrollment concentrated in a few districts versus metropolitan areas with enrollment spread evenly over many districts biased against finding positive effects of competition among districts. Formally, the observed degree of choice available among public school districts may be simultaneously determined with the school quality experienced by the typical student.

To obtain unbiased estimates, we need to identify geographic or historical factors that increase a metropolitan area's tendency to contain many small, independent school districts. We need instrumental variables related to the demand for independent school districts but unrelated to contemporary public school quality. I use the fact that metropolitan areas with more streams had more natural barriers and boundaries that increased students' travel time to school and caused the initial school district lines to be drawn up into smaller districts.¹⁰

This estimation strategy allows me to control for a wide range of background variables that might also influence schools or students. For instance, I control for the effect of household income, parents' educational attainment, family size, family composition (for example, single-parent households), race, region, and metropolitan area size, as well as for the local population's income, racial composition, poverty, educational attainment, and urbanness. Also, because I have good measures of self-segregation by school and school district (for racial, ethnic, and income segregation), I can also differentiate the effects of choice among school districts on self-segregation from those on student achievement or school efficiency.¹¹

My best estimates of the effects of competition among public school districts are gauged in terms of a one-standard-deviation increase in the Herfindahl index. This

corresponds to a substantial increase in the degree of choice among districts; for instance, it is the difference between having three and thirteen equal-sized districts or the difference between having four and a very large number (say, one hundred) of equal-sized districts. An increase of one standard deviation in the degree of choice among districts causes a small (and statistically significant) improvement in student achievement. Students' reading and math scores improve by about 2 percentile points, for instance. However, an increase of one standard deviation in choice among districts causes a large improvement in schools' efficiency. This increase occurs because the small improvement in student achievement takes place even though schools lower their per pupil costs by 17 percent when they face a standard deviation increase in choice. What is striking is the opposite sign of these effects: an increase in choice improves student achievement even while it accomplishes substantial cost savings. The implications for schools' productivity (the ratio of student achievement to dollars spent) are powerful.

What about the effects of competition among districts on the segregation of students? These effects turn out to be insignificant for a reason that may not occur to us at first glance. The degree of racial, ethnic, and income segregation that a student experiences is related to the degree of choice among *schools* in a metropolitan area, but not to the degree of choice among *districts*. (In fact, the point estimates have the "wrong" sign for the latter relationship.) In other words, students are just as segregated in schools in metropolitan areas that contain only a few districts as they are in metropolitan areas that contain many districts. Households sort themselves into neighborhoods inside districts; neighborhoods and schools are small enough relative to districts so that district boundaries have little effect on segregation. This result demonstrates how important it is to compare realistic alternatives. The realistic alternative to a metropolitan area with a high degree of choice among districts is not a metropolitan area in which all schools are perfectly desegregated and every student is exposed to similar peers. The realistic alternative is a metropolitan area with a low degree of choice among districts and schools that exhibit substantial segregation.

EVIDENCE OF THE EFFECTS OF PRIVATE SCHOOL COMPETITION

Choice among public school districts has several other effects worth noting. First, choice among districts and choice between public and private schools are substitutes for one another. An increase of one standard deviation in the degree of choice among districts lowers the share of children who attend private schools by about 1 percentage point (on a base of about 12 percentage points). When parents have more choice within the public sector, they are more likely to be satisfied by their public options, and they are less likely to choose a private option.

Second, when parents have more choice among districts, they tend to be more involved in their children's schooling.¹² For instance, an increase of one standard deviation in the degree of choice causes one out of every three parents to visit the school in the course of a year and causes school administrators to say that parents have a more significant influence on school policy.¹³ Furthermore, parents appear to induce schools to actually pursue the policies that parents *say*, on average, they want in surveys: more challenging curricula, stricter academic requirements, and more structured and discipline-oriented environments. For instance, one standard deviation in the degree of choice in a metropolitan area raises by 8 percent the probability that a school's *regular* mathematics sequence ends in a twelfth-grade course that contains at least some calculus.¹⁴

Finally, the beneficial effects of choice among districts on schools' productivity depend on districts' having a significant degree of fiscal independence. In states such as California, where districts depend almost entirely on state per student allocations for their budgets, the positive effects of choice on student achievement and cost savings are reduced by about one-half. This is probably because successful schools are not rewarded through the property tax/budget process for their efforts to improve achievement or reduce costs. This result has implications for analyses of reforms: researchers should consider that reforms do not always give participating schools sufficient fiscal independence to allow them to benefit financially from their own success.

To determine the effects of private school competition on public schools and public school students, we can also compare areas with and without substantial private school enrollment. However, low-quality public schools raise the demand for private schools as substitutes for public schools. Therefore, such simple comparisons would confound the effect of greater private school competitiveness with the increased demand for private schools where public schools are poor in quality. Formally, private school enrollment is likely to be endogenous to (partly caused by) public school quality, and this endogeneity would lead simple estimates to be biased toward finding that private school competition had negative effects on public schools.

To obtain unbiased estimates, we need to identify factors that increase the supply of private schools in an area and that are unrelated to public school quality. Formally, we need instrumental variables that shift the supply of private schools and that are unrelated to the demand for private schools that is generated by low public school quality. I use the fact that a denomination's private schools have more resources with which to provide tuition subsidies in areas that are densely populated by that denomination. Since religious composition of an area is largely a matter of historical accident, it is not likely to have an independent effect on public school quality. Areas with higher Catholic population shares, for instance, have a larger share of teaching services donated by members of religious orders (worth 30 to 35 percent of costs) and provide a larger share of Catholic school income through donations from the diocese and local households (25 to 50 percent of costs). Thus, denominations' population shares fulfill the conditions for a good instrument: they are positively correlated with the supply of private schools but are likely to be uncorrelated with the part of the demand for private schools that is generated by public school quality. Catholic population shares provide the best instrumental variables not only because school subsidies are a relatively high-priority use of Catholic Church funds, but also because Roman Catholicism is spread across the entire United States rather than concentrated in one state or one region.

Roman Catholicism is also associated with many ethnic groups, unlike some other denominations, which are associated with only one or two ethnic groups.

Note that this estimation strategy allows me to control for a variety of background factors that might be correlated with both the demand for private schools and public school quality (or public school students' performance). For instance, I control for the effect of a household's belonging to a denomination. If *being Catholic*, say, affects a household's demand for public school spending or the achievement of its children, this effect is controlled for (and not confounded with the effect of more or less private school competition). I also control for the effect of certain ethnic group concentrations in an area, for the effect of racial and ethnic homogeneity in an area, for the effect of religious homogeneity in an area, and for the religiosity of an area. Numerous other background factors are controlled for: family income, the share of households in poverty, parents' educational attainment, family size, family composition (single-parent households), urbanness, population density, and region of the country.¹⁵

My best estimates of the effect of more competition from private schools suggest that if private schools in an area receive sufficient resources to subsidize each student's tuition by \$1,000 then the achievement of *public* school students rises. This is true whether the measure of achievement is test scores, ultimate educational attainment, or wages. The effect on mathematics and reading scores is an 8 percentile point improvement. The effect on educational attainment is an 8 percent increase in the probability of graduating from high school and a 12 percent increase in the probability of getting a baccalaureate degree. The effect on wages (for those who work later in life at ages twenty-nine through thirty-seven) is a 12 percent improvement.¹⁶

Interestingly enough, the estimates indicate that competition from private schools does not have a significant effect on public school spending per pupil.¹⁷ This is probably because the two forces described above offset one another. On the one hand, an increased supply of private schools tends to draw into the private school sector many parents who might have supported generous public school

spending if their children had remained in public schools. This phenomenon tends to decrease voter support for public school spending. On the other hand, an increased supply of private schools draws into the private school sector many students who would otherwise have had to be educated at the public's expense. This phenomenon tends to increase public school spending *per pupil*.

What about the effects of private school competition on the self-segregation of students among schools? I will not dwell on these estimates, because their ability to predict the effect of a private school voucher program is limited. The reason is that the estimates are based on private schools that have a religious affiliation, mainly Catholic schools. In contrast, proposed voucher programs often exclude private schools with a religious affiliation and always constrain private schools that accept vouchers to either accept all voucher applicants or to accept some random sample of them.

The estimates do have general applicability in one regard, however: all the self-segregation effects are very small. There are two reasons: First, public schools are *already* quite segregated along lines of race, ethnicity, parents' income, and students' performance. When people imagine the effect of increasing private school availability, they sometimes conjure up a notional public school that is perfectly desegregated. The effects of private school competition on such a notional public school might be dramatic. Even if we could estimate such effects, however, they would be irrelevant since actual public schools do not correspond closely to this ideal. The actual self-segregation effects of traditional private school competition are small simply because a large increase in self-segregation cannot be obtained by sorting out an already segregated public school. The second reason that self-segregation effects are small is that an increase in private school competition typically allows self-segregation in public schools to increase slightly while self-segregation in private schools decreases slightly. These effects tend to offset one another.

My best estimates suggest that if private schools in an area receive enough resources to subsidize tuition by \$1,000, segregation along lines of race, ethnicity, income, and student performance decreases at private schools by

small, but statistically significant, amounts. At the same time, segregation along these lines changes at public schools by amounts that have positive point estimates but are statistically not different from zero.¹⁸

Finally, note that both private school competition and competition among public schools tend to hold down input costs. Specifically, both types of competition constrain the salary increases that teachers' unions gain for their members (the union wage premium of 12 percent is reduced by about one-third for a standard-deviation increase in competition among districts and by about one-half for a \$1,000 subsidy for private schools).¹⁹ This result parallels a standard result from private industry: increased competition in the market for a product (in this case, the market offering schooling to students) tends to decrease the wage premia earned by unionized workers and other inputs that are provided by suppliers with market power.

EVIDENCE ON THE EFFECTS OF INTRADISTRICT CHOICE PROGRAMS

Intradistrict choice has been utilized by a number of large school districts for some time. The least dramatic forms of intradistrict choice are magnet or alternative schools, to which a student typically applies because of a preference for an alternative curriculum or schooling environment. In the more dramatic forms of intradistrict choice (Manhattan's District 4 or Cambridge, Massachusetts), every student must actively express a preference for a school. Intradistrict choice shares some features of the two traditional forms of school choice discussed above. In particular, the fact that parents and students make an active choice is likely to make them more committed and involved in schooling. However, intradistrict choice programs rarely give schools the degree of fiscal or curricular autonomy enjoyed by independent school districts or private schools. It is important to recognize that a district that gives fiscal or curricular autonomy to a school in a given year has not given the school long-term autonomy unless the district can bind itself to not revoke that autonomy. Such binding often proves to be politically impossible. For instance, intradistrict choice programs sometimes exhibit long-term fiscal incentives that are perverse because the district cannot,

ex post, resist taking money from successful schools and giving it to unsuccessful schools.

The evidence on intradistrict choice is at an exploratory stage. My own work demonstrates only that simple estimates (comparing districts that have intradistrict choice with districts that do not) are badly biased.²⁰ The bias arises because districts do not randomly enact intradistrict choice programs. These programs are usually associated with the hiring of a superintendent who is given a free hand to "turn around" a district that has recently experienced sharp decreases in student achievement. It is difficult to create a control group of schools that can be compared effectively with this type of school. Even before-and-after studies do not enable us to disentangle the effects of intradistrict choice from the effects of getting a new superintendent who is paid more and given greater latitude than previous administrators.²¹

LESSONS FOR REFORM FROM TRADITIONAL SCHOOL CHOICE

The evidence on the effects of traditional school choice gives us several lessons that are helpful for analyzing reforms.

- First, public schools can and do react to competition by improving the schooling they offer and by reducing costs. They are not passive organizations that allow their students and budgets to be withdrawn without responding. Realistic increases in the competition they face produce significant improvements in students' test scores, educational attainment, and wages.
- Second, public schools' responses do not depend just on whether they lose students; the responses also depend on the fiscal rewards and penalties attached to gaining or losing students. When competition has little fiscal implication, a public school is less likely to react. When cost competition is weakened by a large price wedge (like that between public and private schools), public schools reduce costs less than they do when cost competition is on a more level playing field (like that between two similar public school districts).
- Third, the segregation effects of increasing school choice via reforms are likely to be small because schools in the United States (not merely districts) are already quite segregated. To predict accurately the

effects of reforms on segregation, one must consider a realistic alternative, not an idealized public school with perfect desegregation.

- Fourth, parents who have greater choice are more involved in their children's schooling. Parents' influence on school policy, which is greater when choice is greater, will reflect, on average, their stated preferences for tougher curricula and stricter school atmospheres. Note, however, that greater choice is also likely to make schools more diverse through parental influence because like-minded parents will be better able to group together in schools. (I have no evidence on this last point.)
- Finally, different types of school choice substitute for one another to a limited degree.

Given these lessons, what other pieces of information do we need in order to analyze school choice reforms? Three information deficiencies stand out. Since we know that the fiscal impact of a choice program is an important determinant of the program's effect on schools, the financial arrangements of charter school programs, open enrollment programs, and vouchers will be key determinants of their effects. These financial arrangements often receive little thought, and they are chosen more for convenience and political reasons than because they generate good financial incentives. States that want to avoid perverse financial incentives should consider financial arrangements that purposely mimic the fiscal impacts of the two traditional forms of school choice. In order to estimate the effects of more dramatic fiscal incentives, we will need to observe actual choice reforms that carry a variety of financial arrangements.

The second information deficiency pertains to the long-term sustainability of reforms. All three of the reforms discussed create schools or programs that have less long-term autonomy than do the schools that compete in the two traditional forms of school choice. Public school districts have indefinite lifetimes and will not have difficulty raising tax revenues as long as parents want to send their children to the schools. Private schools have similarly

indefinite lifetimes and can raise tuition revenue so long as they attract parents. While some charter school laws are written to give a high degree of fiscal autonomy to charter schools, all charter schools must get their charters renewed by the state (at least) and depend on other organizations to decide their per pupil payments. It remains to be seen whether charters and per pupil payments are politically maintainable when and if charter schools become successful competitors for the revenues and students of public school districts. Most open enrollment programs have even less inherent political sustainability. These programs, at least as written thus far, require the ongoing cooperation of local public school districts. (The receiving district must almost always cooperate voluntarily, although involuntary cooperation is sometimes exacted from the sending district.) The voucher programs passed to date depend on the sufferance of the sending district, but some proposed programs make the vouchers less dependent on that district. Careful analyses of district-level and state-level politics will be necessary to predict the long-term sustainability of all three reforms.

Finally, traditional school choice gives us only limited information about the supply response we can expect from private schools under a voucher program or from charter schools. Supply responses are estimated in the analyses of choice among public schools and choice between public and private schools. (For instance, giving private schools additional resources that are equivalent to a \$1,000 tuition subsidy creates a 4.1 percentage point increase in Catholic school enrollment, from a base of about 10 percent.) However, proposed charter school programs and voucher programs sometimes take us beyond the range where extrapolation from traditional school choice results is reasonable. A voucher of \$3,500 available to *all* poor students, for instance, would produce a long-term supply response that would be difficult to predict since the availability and long-term horizon exceed those of current voucher programs (like Milwaukee's), and the voucher amount exceeds those of most current private school subsidies.

ENDNOTES

1. Some people unfortunately associate the word “segregation” exclusively with racial segregation. I am using it to describe segregation of students along a number of lines. It could also be described as student sorting, and it encompasses a variety of phenomena, including segregation by ability, sometimes called “cream skimming” or “cherry picking.”

2. See Hoxby (1996a, 1996b, 1996c, 1996d, 1997a, 1997b). Copies of unpublished papers can be obtained by sending a written or electronic request to the author.

3. It is important to note that the fiscal reward process works through the residential decisions of *marginal* home buyers. If marginal home buyers choose to locate in other districts because district X is a poor or inefficient provider of schooling, then all home prices in district X fall. There is no need for all households to relocate for all houses’ prices to affect the districts’ fiscal rewards. See Hoxby (1996a) for details.

4. In practice, however, some states’ charter school laws allow the schools very little administrative or fiscal autonomy. For instance, a charter school has little administrative autonomy if it is automatically subject to all clauses of the local teachers’ unions’ collective bargaining agreements. Similarly, a charter school may have little fiscal autonomy if its tuition payments depend completely on the per pupil spending of the local school district (regardless of the charter school’s own success).

5. Public schools must admit all students in their attendance district. Charter schools and open enrollment schools must admit a random sample of students who are eligible and interested in attending.

6. There are and have been some minor public subsidies for private school expenses, including small tuition tax deductions and credits. Some states also require local public districts to provide certain textbooks and bus transportation to private school students.

7. Although tuition understates the true cost of private schooling, private schooling does cost significantly less than public schooling on average. Over the entire period from 1976 to the present, per pupil costs in private schools have always been between 50 and 60 percent of contemporary per pupil costs in public schools.

8. For this section, see Hoxby (1997a).

9. A Herfindahl index based on enrollment shares is as follows. Suppose a metropolitan area has J school districts, which we index by $j=1, \dots, J$.

Suppose each school district has a share, s_j , of total metropolitan area enrollment. Then, the inverse Herfindahl index is

$$-\sum_{j=1}^J s_j^2 .$$

When there is no choice in a metropolitan area because there is only one public school district, the inverse index is equal to -1. As more districts are added and as enrollment is spread more evenly over those districts, the inverse index gets closer to zero.

10. This typically took place about the time of Anglo-American settlement, which varies with the area of the country. Many of the original petitions for district boundaries cite streams as a reason for not extending the district lines. Streams are by far the most common natural boundary for school districts. Note, however, that many of the streams that are preserved in boundaries are small and have never had industrial importance. Today, many of the boundary streams are of negligible importance in travel.

11. The equations estimated can be summarized as follows. The main equation to be estimated is of the form

$$y_{ik} = aH_k + X_{ik}\beta + X_k\zeta\delta + \varepsilon_k + \varepsilon_{ik} ,$$

where y is an outcome such as a student’s test score or a school’s per pupil spending, i indexes students or schools (depending on the outcome), k indexes the metropolitan area, H is the inverse Herfindahl index that measures the degree of choice among public school districts, X_{ik} is a vector of background variables that describe the student or school (for instance, the race and gender of the student or the homogeneity of household incomes for students who attend the school), and X_k is a vector of background variables that describe the metropolitan area (for instance, its racial composition and size). The two-tiered error structure adjusts the standard errors for the fact that the degree of choice varies only at the metropolitan area level.

There is also an implied first-stage equation that estimates the effect of streams on the concentration of public school districts in the metropolitan area:

$$H_k = S_k\gamma + \bar{X}_{ik}\kappa + X_k\lambda + v_{ik} ,$$

where H_k , \bar{X}_{ik} , and X_k are as above (except that \bar{X}_{ik} is effectively averaged for the area), and S_k is a vector of variables that measure the prevalence of large and small streams in the metropolitan area.

12. See Hoxby (1996d).

13. Specifically, the measure of parental influence over school policy rises by two-thirds of one standard deviation.

ENDNOTES (*Continued*)

14. Interestingly, an increase in the degree of choice encourages grade inflation, which I measure by comparing students' course grades with their performance on national standardized exams in the same subjects. This finding suggests that although parents want their children to be exposed to harder "real" curricula, parents are loath to set higher "nominal" standards for their children—perhaps because local grade deflation might be misinterpreted by colleges in the admissions process.

15. The equations estimated can be summarized as follows. The main equation to be estimated is of the form

$$y_{ik} = \mu V_k + X_{ik}\nu + X_k\pi + \upsilon_k + \upsilon_{ik} ,$$

where y is an outcome such as a student's wage or a school's per pupil spending, i indexes students or schools (depending on the outcome), k indexes the area (metropolitan areas and counties, depending on the urbanness), V is the average tuition subsidy offered by private schools in area k , X_{ik} is a vector of background variables that describe the student or school (for instance, the student's own religion or the racial homogeneity of the school), and X_k is a vector of background variables that describe the area (for instance, its income composition or religiosity). The two-tiered error structure adjusts the standard errors for the fact that average tuition subsidies vary only at the area level.

There is also an implied first-stage equation that estimates the effect of denominations' population shares on the tuition subsidies offered by private schools:

$$V_k = D_k\rho + \bar{X}_{ik}\theta + X_k\tau + \omega_{ik} ,$$

where V_k , \bar{X}_{ik} , and X_k are as above (except that \bar{X}_{ik} is effectively averaged for the area), and D_k is a vector of population shares of denominations $m=1, \dots, m$ in area k .

16. These are instrumental variable estimates of the effect of a \$1,000 subsidy for private school tuition where the equations are as in endnote 15. The coefficient estimates and their standard errors are, respectively, 7.9 (3.5); 2.2 (1.0); 3.3 (1.1); 12 (5.7). See Hoxby (1996b, 1997b).

17. See Hoxby (1996b).

18. Statistical significance here refers to asymptotic statistical significance at the 10 percent level. Income segregation is measured using students' free-lunch eligibility. See Hoxby (1997b).

19. See Hoxby (1996c).

20. See Hoxby (1996d).

21. In addition, before-and-after studies suffer from bias produced by "Ashenfelter's dip." This dip is simply the phenomenon that treatment (in this case, intradistrict choice) is frequently assigned to individuals (in this case, school districts) who have recently experienced a negative departure from their own history. Since individuals and districts would typically experience mean reversion anyway (and return to their historic paths), simple before-and-after studies tend to exaggerate the positive effect of treatment.

REFERENCES

- Hoxby, Caroline.* 1996a. "Are Efficiency and Equity in School Finance Substitutes or Complements?" *JOURNAL OF ECONOMIC PERSPECTIVES* 10, no. 4 (fall).
- . 1996b. "The Effects of Private School Vouchers on Schools and Students." In H. Ladd, ed., *HOLDING SCHOOLS ACCOUNTABLE*. Washington, D.C.: Brookings Institution.
- . 1996c. "How Teachers' Unions Affect Education Production." *QUARTERLY JOURNAL OF ECONOMICS* 111, no. 3 (August).
- . 1996d. "When Parents Can Choose, What Do They Choose?" Unpublished paper, Harvard University.
- . 1997a. "Does Competition among Public Schools Benefit Students and Taxpayers?" Revision of NBER Working Paper no. 4979 (1994).
- . 1997b. "Do Private Schools Provide Competition for Public Schools?" Revision of NBER Working Paper no. 4978 (1994).
- Tucker, A., and W. Lauber.* 1995. *SCHOOL CHOICE PROGRAMS: WHAT'S HAPPENING IN THE STATES*. Washington, D.C.: Heritage Foundation.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

Schools and Student Achievement: More Evidence from the Milwaukee Parental Choice Program

Cecilia Elena Rouse

Many states are considering programs that would provide vouchers for (low-income) children to attend private schools because policymakers believe that traditional reforms—such as reducing class sizes—will not fix an educational system that is “broken.” Advocates of vouchers argue that teachers’ unions and bloated bureaucracies impede such reforms from reaching the classroom and increasing student achievement. Furthermore, because children are required to attend their neighborhood school, the system has no incentive to change. Wealthier parents can voice dissatisfaction with their residential school by moving to another neighborhood or enrolling their children in a private school; however, poorer—particularly inner-city—parents cannot. Vouchers would, at a minimum, provide disadvantaged children with more educational options. If the students also received a better education in

the private schools, the program might offer a cost-effective way to improve student achievement, at least for those students who use the vouchers.

In 1990, Wisconsin became the first state in the nation to implement a publicly funded school voucher program. The Milwaukee Parental Choice Program provides a voucher, worth approximately \$4,373 in 1996-97, to low-income students to attend nonsectarian private schools. The program began with seven private schools, although by 1996 the number had risen to twenty.¹ At this time, religious schools are not permitted to participate in the program.² The participating private schools offer a variety of educational approaches, including Montessori and Waldorf, as well as bilingual and African-American cultural emphases. Although the tuition charged by many of the “choice” schools is quite low (ranging from less than \$200 to about \$4,000), actual expenditures per pupil are generally higher (on the order of \$4,000 to \$5,000 per pupil in 1996-97).³ The balance of the revenues comes from grants, donations, and fund-raising by parents. In

Cecilia Elena Rouse is an assistant professor of economics and public affairs at Princeton University and a faculty research fellow of the National Bureau of Economic Research.

addition, because the schools are nonsectarian, many also receive Title I funding from the federal government.

Because the parental choice program is targeted to the most disadvantaged public school students, only students whose family income is at or below 1.75 times the national poverty line are eligible. In principle, the student in a family of three with a family income of approximately \$21,000 is eligible to apply; in practice, the mean family income of applicants is approximately \$12,300. Choice applicants are considerably more disadvantaged than the average student in the Milwaukee public schools (whose family income is \$24,000); they are also more likely to be minority and have lower preapplication math and reading test scores. However, the parental education of choice applicants is comparable to that of nonapplicants.

Some argue that an unrestricted voucher program would improve the schooling of all children. In the most unrestricted program, all (or a substantial percentage) of the students in the public schools would be eligible to attend a private school. Since state funding would be tied to student enrollments, public schools would have to compete for students, as in the marketplace, which would give the schools an incentive to improve. If such an unrestricted voucher program were successful, the academic outcomes of students in public and private schools would equalize over the long run. While such effects are theoretically possible, the Milwaukee Parental Choice Program is too small to provide insight into the potential student achievement benefits of an unrestricted voucher program.⁴ It cannot show whether providing vouchers would also improve the schooling of students who remain in the public schools. An analysis of the Milwaukee Parental Choice Program can, however, indicate whether the private schools participating in the program (the choice schools) are "better" than the public schools in Milwaukee.

In this paper, I review the three existing studies of the effects of the choice schools on student achievement. Two of the studies report significant gains in math for the choice students and two report no significant effects in reading. I also extend the analysis to compare the achievement of students in the choice schools with that of students in three different types of public schools: regular

attendance area schools, citywide (or magnet) schools, and attendance area schools with small class sizes and supplemental funding from the state of Wisconsin ("P-5" schools). The results suggest that students in P-5 schools have math test score gains similar to those in the choice schools, and that students in the P-5 schools outperform students in the choice schools in reading. In contrast, students in the citywide schools score no differently than students in the regular attendance area schools in both math and reading. Given that the pupil-teacher ratios in the P-5 and choice schools are significantly smaller than those in the other public schools, one *potential* explanation for these results is that students perform well in schools with smaller class sizes.

EXISTING STUDIES OF THE ACHIEVEMENT EFFECTS OF THE CHOICE PROGRAM

Three studies to date have evaluated the achievement effects of the Milwaukee Parental Choice Program. The first, conducted by Witte, Sterr, and Thorn (1995), concludes that there were no relative achievement gains among the choice students (see also Witte [1997]). The second, by Greene, Peterson, and Du (1997), finds that the choice students made statistically significant test score gains in both reading and math by their third and fourth years in the program. The third study, by Rouse (forthcoming), reports that the students selected to attend a choice school experienced significantly faster gains in math scores, but showed no differential gains in reading. To understand why these three studies generated conflicting results, one must consider two aspects of the evaluations: the selection of the control, or comparison, group and the method of controlling for family background and student ability.

SELECTION OF THE CONTROL, OR COMPARISON, GROUP

Ideally, to establish whether choice schools are better than the Milwaukee public schools, one must ascertain whether students who attended the choice schools had higher achievement gains than they *would have had* if they had attended a Milwaukee public school. Because this counterfactual is impossible to obtain, one must instead identify a group of students who did not attend a choice school; their

test scores provide the yardstick against which to measure the effect of the program. This group is called a control, or comparison, group.⁵ The best control group is constructed using a randomized experiment. In this social experiment, children are randomly assigned to attend a choice school (the treatment group), while others are assigned to attend public schools (the control group). After some period of time, one would compare outcomes—such as test scores, high school graduation rates, or labor market success—for the treatment and control groups. Since, on average, the only difference between the groups would be their initial assignment—which was randomly determined—any differences in outcomes could be attributed to the type of school attended.

Such an experiment, however, was not implemented in Milwaukee (nor anywhere else), forcing researchers to devise statistical methods that attempt to mimic a randomized experiment. One cannot simply compare the achievement of students in choice schools with that of a comparison group of students in the Milwaukee public schools. In Milwaukee, this simple comparison would likely show that students enrolled in choice schools fare no better than students in the Milwaukee public schools.⁶ One might be tempted to conclude that the choice schools are no different than the public schools. However, such an interpretation might be misleading. Students who qualify for the parental choice program come from disadvantaged families. As a result, they generally score lower on standardized tests than wealthier, more advantaged students and would likely have continued to do so even if they had remained in the public schools. One would attribute the test score results to the schools when the results may, in fact, be due to the characteristics of the students. To estimate the true effect of the choice schools, one must control for family background (such as family income and parental education) and student ability. The goal is to control for all individual characteristics that are correlated with attending the choice school and to explain the higher test scores in such a way that the only difference between the two groups of students is enrollment in a choice school. In general, the more similar the two groups are to begin with, the more credible the evaluation of the program will be.

The choice of a control, or comparison, group is one area in which the existing analyses of the Milwaukee voucher program differ. Greene, Peterson, and Du (1997) compare the test scores of choice students with those of the group of students who applied to the program but were not accepted (the “unsuccessful applicants”); Witte, Sterr, and Thorn (1995) compare choice students with a random sample of students from the Milwaukee public schools; and Rouse (forthcoming) compares selected choice students with both the unsuccessful applicants and the students in the Milwaukee public schools. There are advantages and disadvantages to both control/comparison groups.

The unsuccessful applicants are an appealing control group because all of these students were interested in attending a choice school. Therefore, the unsuccessful applicants likely have parents who are similarly motivated to the parents of the successful applicants. In addition, the parents of all applicants must expect that their children will be well served in the program, which may not be true for the children who did not apply. There are problems with using the unsuccessful applicants as a control group, however. The first is that since the parents of all applicants were interested in enrolling their children in a private school, the parents of the unsuccessful applicants may have been more likely to enroll their children in a private school outside of the choice program, rather than re-enrolling them in a Milwaukee public school. This decision was made easier by a parallel, privately funded program—Partners for Advancing Values in Education (PAVE)—that provided scholarships to students interested in attending (primarily) Catholic schools. If post-application data on these students were available, this would not be a problem. However, the data do not track students who enrolled in either a public school outside of the Milwaukee public school system or a nonchoice private school. The second problem is that the sample sizes are extremely small. By the fourth year of the program, there were fewer than forty unsuccessful applicants to use in evaluating the program, which makes estimated effects of the program sensitive to unusually high or low test scores (Witte 1997).

One can also compare the achievement of students in the choice schools with that of a random sample of

students from the Milwaukee public schools. This comparison group yields a much larger sample and is, perhaps, less subject to nonrandom attrition (after all, these students were ostensibly not interested in leaving the Milwaukee public schools). At the same time, the random sample of students from the Milwaukee public schools may have refrained from applying to the parental choice program because they thought it would not serve them well, or because their parents are less motivated or involved, which would lead to an overstatement of the achievement effects of the program. As a result, using this comparison group requires a statistical strategy that adequately controls for student characteristics.

METHOD OF CONTROLLING FOR STUDENT CHARACTERISTICS AND FAMILY BACKGROUND

The second area in which the existing analyses of the Milwaukee program differ is the method of controlling for family background and student ability. Greene, Peterson, and Du (1997) control for “application lotteries”; Witte, Sterr, and Thorn (1995) control for the student’s prior test scores; and Rouse (forthcoming) controls for “individual fixed effects.” Again, each methodology has advantages and disadvantages.

Consider first the strategy employed by Greene, Peterson, and Du. The choice schools are not allowed to discriminate in admitting students, which is interpreted to mean that if more students apply for the school than there are seats available, the students are randomly selected from among the applicants. If a choice school is not oversubscribed, it is required to take all who apply, with only a few exclusions. Therefore, in each school in which students are randomly selected (through an application lottery), a mini-randomized experiment is conducted. If the schools truly select the students at random, then, on average, the only difference between the successful and unsuccessful applicants is whether they have been randomly selected. As a result, in theory, one could simply compare the outcomes of successful applicants with the outcomes of unsuccessful applicants and attribute the difference to whether the students were selected to attend a choice school. Moreover, because selection was random (that is, not related to student

ability or parental background), one need not control for individual characteristics.⁷ One can also combine all of these mini-experiments and control for variables indicating the application lottery in which each student participated.⁸ (Naturally, this strategy requires using the unsuccessful applicants as a control, or comparison, group.) The primary advantage of using the unsuccessful applicants as a control group and controlling for application lotteries is that, if selection is truly random, this strategy should uncover the true effect of the parental choice program on student test scores using a method that closely resembles a randomized experiment, at least in theory.

In practice, this strategy has some disadvantages. First, the data do not contain information on the actual school(s) to which a student applied. As a result, one cannot recover the actual application lotteries. Greene, Peterson, and Du have devised a creative solution to this problem, but it is not clear how close their imputation comes to the actual lotteries.⁹ A second disadvantage is that even if the lotteries are truly random and the imputation reasonably mimics them, it appears that the motivated unsuccessful applicants were more likely to attend another private school—one outside of the choice program (Rouse forthcoming; Witte 1997). As a result, by not controlling for family background, one may overstate the effectiveness of the program.

There are also advantages and disadvantages to controlling for prior test scores—the methodology implemented by Witte, Sterr, and Thorn. On the one hand, controlling for these scores has the advantage of accounting for student ability that changes over time, rather than controlling for characteristics at a fixed point in time. In addition, this methodology allows one to develop a dynamic model of test score growth in which a child’s test score this year is a direct function of his or her test score last year. On the other hand, test scores may not be a good measure of ability (even ability at a fixed point in time). Moreover, the strategy may not be appropriate when applied to data on students who have been enrolled in a choice school for several years (Rouse forthcoming). Finally, one can only include students who have prior test scores in the analysis, which is a potential problem in Milwaukee, where the majority of students are not tested each year.

The strategy implemented by Rouse (forthcoming) controls for all student characteristics (both observed and unobserved) that do not change over time (that is, they are fixed, or time-invariant). These characteristics include more motivated parents, parental education, and innate student ability. The methodology is referred to as controlling for individual fixed effects. This strategy requires fewer assumptions than one using application lotteries and allows for larger samples than one controlling for prior test scores. Its primary disadvantage, however, is that it does not control for time-varying student characteristics.

To understand this strategy (which I employ in the rest of the paper), consider two students: Student A, who enrolled in a choice school, and Student B, who did not. The diagram depicts two possible test score trajectories for the two students *before Student A enrolled in the choice program* (see box).¹⁰ Suppose that the prechoice test scores of Student A and Student B evolve as shown in the left portion of the diagram. Here, Student A scored higher than Student B each year before Student A enrolled in the program. This may reflect the fact that Student A was more “able” than Student B, and one would not want to

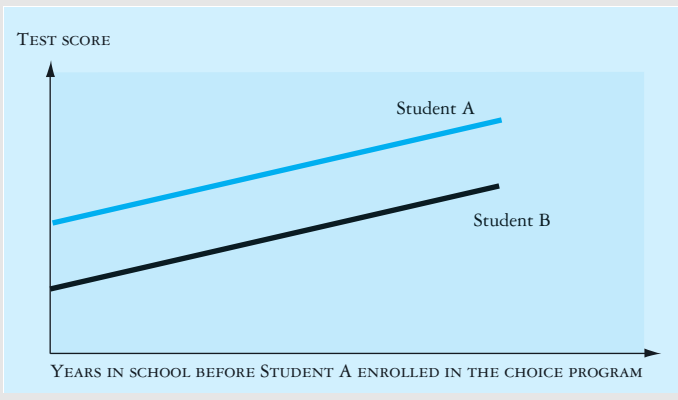
attribute the test score difference to the choice schools since it existed even before Student A enrolled in the choice program and it would likely have continued to exist even if Student A had remained in a Milwaukee public school. Fortunately, in this case the fixed-effects analysis will uncover the true (unbiased, in statistical terms) effect of the choice schools on student achievement.

The fixed-effects analysis will, however, lead to an overstatement of the program’s effects if Student A had faster test score gains than Student B before Student A enrolled in the choice program. In this case (shown in the right portion of the diagram), the fixed-effects analysis will attribute the faster achievement growth to the choice program when, in reality, students in the choice program would have had faster test score growth even if they had remained in the Milwaukee public schools. To assess whether this potential problem likely explains the entire estimated program effect, I analyzed the preapplication test score trajectories of students in the choice program and those in the Milwaukee public schools. This exercise indicated that the results obtained using individual fixed-effects estimates are probably not overstated.

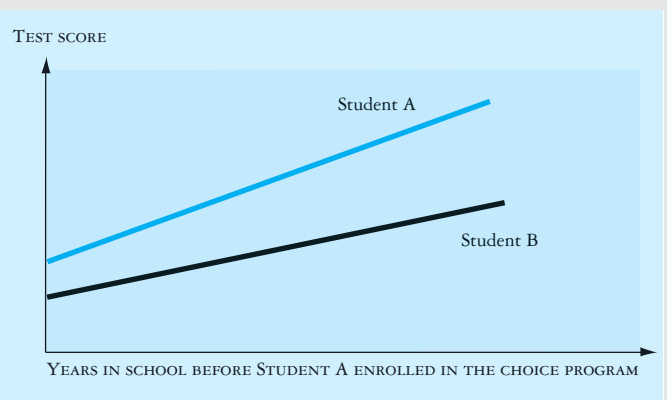
UNDERSTANDING INDIVIDUAL FIXED-EFFECTS ESTIMATES

Student A is enrolled in a choice school and Student B is enrolled in a Milwaukee public school. Consider their test scores before Student A enrolled in the choice program:

Individual fixed-effects estimates will generate the true effect if both Student A and Student B had the same growth in test scores before Student A enrolled in the choice program.



Individual fixed-effects estimates will overstate the effect if Student A had faster test score gains even before he or she enrolled in the choice program.



ARE THE PRIVATE SCHOOLS IN THE CHOICE PROGRAM “BETTER” THAN THE MILWAUKEE PUBLIC SCHOOLS?

COMPARING CHOICE SCHOOLS WITH ALL MILWAUKEE PUBLIC SCHOOLS

Chart 1 compares the test scores of students selected to attend a choice school with those of both the unsuccessful applicants and the random sample of students from the Milwaukee public schools, controlling for individual fixed effects.¹¹ Note that I use the test scores of those *selected* to attend a choice school, whether or not the student ever enrolled in a choice school or eventually returned to the Milwaukee public schools. I show these results for two reasons. First, making vouchers available is the only policy instrument open to policymakers. If the state of Wisconsin decides to provide educational vouchers to all low-income

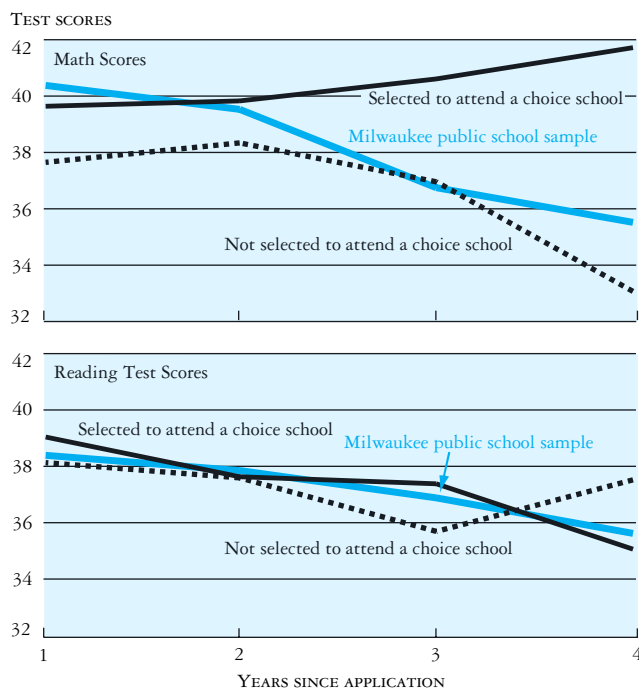
students, not all will take advantage of the program and not all who enroll will remain in it. In the extreme case in which no students actually use the vouchers, even if the choice schools are much better at educating children than the public schools are, there will be no achievement gains from the program. Thus, comparing the test scores of students who are selected (whether or not they actually are enrolled in a choice school) reflects the overall potential gains from offering the vouchers. Second, students who leave the choice schools may do so because they are not flourishing there. In this case, an analysis that compares the test scores of students who remain enrolled in a choice school may overstate the true effect of the program.¹²

The top panel of Chart 1 shows that students selected for the choice program made yearly gains in math achievement, particularly beginning in the second year after application. It also reveals that both the unsuccessful applicants and the students in the Milwaukee public school sample experienced large declines in their math test scores in the third and fourth years. The bottom panel shows the trends for reading scores. It is clear that there are no discernible differences in the reading test scores between the three groups.

Given that the trends for the unsuccessful applicants and the students in the random Milwaukee public schools sample are similar, Chart 1 shows that any differences between the three existing analyses do not hinge on the selection of a control, or comparison, group (provided that family background is adequately controlled for). In addition, the math results in the chart are consistent with those reported by Greene, Peterson, and Du (1997), and the reading results accord with those reported by Witte, Sterr, and Thorn (1995). The fact that the math results agree with those reported by Greene, Peterson, and Du indicates that *in these data*, if one adequately controls for student characteristics, it does not make a large difference whether one defines choice students as those who are *selected* to attend a choice school or as those who are *enrolled* in a choice school. In contrast, the reading results conflict with those reported by Greene, Peterson, and Du, largely because the authors’ results disappear when one includes individual fixed effects. The math results conflict with those reported by

Chart 1

ESTIMATES OF MATH AND READING TEST SCORES FOR STUDENTS SELECTED TO ATTEND A CHOICE SCHOOL, APPLICANTS NOT SELECTED, AND STUDENTS IN THE MILWAUKEE PUBLIC SCHOOLS



Source: Rouse (forthcoming).

Note: The estimates control for individual fixed effects (for example, they are corrected for ability and family background).

Witte, Sterr, and Thorn because of differences in our specifications and samples.¹³

It is also worth noting that these data are far from ideal for an evaluation of the choice program. The fact that students who were not enrolled in either a choice school or a Milwaukee public school were not included in the data leads to concerns about nonrandom sample attrition. In addition, because of changes in the tests administered in the public schools, some data are imputed.¹⁴ I continue to estimate results similar to those presented in Chart 1 when I attempt to control for both sample attrition and data imputations. Nevertheless, statistical techniques cannot substitute for better data, so these data deficiencies should be kept in mind when interpreting the results.

COMPARING CHOICE SCHOOLS WITH DIFFERENT TYPES OF PUBLIC SCHOOLS

Other studies have also found that private schools perform better than public schools (see, for example, Coleman, Hoffer, and Kilgore [1982a, 1982b], Evans and Schwab [1995], Neal [1997], and Sander [1996]). Many attribute the observed superiority of private schools to the fact that these schools compete for students.¹⁵ However, few have attempted to look within the “black box” of private school success to understand why the schools may be successful. Those who *have* looked point to differences in homework, curriculum, decentralized governance, and social integration (Bryk, Lee, and Holland 1993; Coleman, Hoffer, and Kilgore 1982a; Coleman and Hoffer 1987). I attempt to look more closely at the apparent Milwaukee private-public school differences in achievement by focusing more intensely on the public schools.¹⁶

The Milwaukee public school district consists of approximately 145 schools. The district operates a controlled choice program in which first-time students in Milwaukee’s public schools, students who reach the top grade of their school, and students desiring to transfer from their attendance area school are required to select three schools in which they would like to enroll. If a school is oversubscribed, selection is based on a random lottery with preference given to children attending the feeder schools, those with siblings already enrolled in the school, and those living in the attendance area or nearby (Milwaukee Public Schools 1997).

Within the district there are approximately thirty citywide (or magnet) schools, which were created in the 1970s to facilitate desegregation. Many of these schools are specialized, offering foreign language immersion, gifted and talented and performing arts instruction, and Montessori, Waldorf, and Global Learning educational approaches. Approximately 22 percent of the total Milwaukee public school enrollment is in citywide schools.¹⁷ Many researchers (for example, Archbald [1995]) hypothesize that citywide schools should be better than regular attendance area schools because citywide schools compete for students (at least within the district). In Milwaukee, this competitive effect may be muted, however, because although the citywide schools are designed to accommodate students from all over Milwaukee, many of them allocate over half of their available seats to children who live close to the school (Milwaukee Public Schools).

Finally, a group of fourteen schools (known as “Project Rise Schools”) whose students are predominately minority and extremely disadvantaged were exempted from desegregation. Instead, they were provided with extra funding from the state. Today, these fourteen schools, along with about seven others, participate in the Preschool to Grade 5 Grant Program, and are known as P-5 schools;¹⁸ they enroll about 15 percent of the total public school students and 25 percent of the elementary school students. This program provides supplemental state grants to schools with high proportions of economically disadvantaged and low-achieving students. In theory, eligible schools are required to maintain pupil-teacher ratios of under twenty-five to one, institute annual testing in basic skills, identify students needing remedial education, increase parental involvement, provide in-service training, and conduct staff evaluations (Clancy, Toulmin, and Bukolt 1995). In practice, the schools primarily comply with the small class size requirement. In 1993-94, Wisconsin allocated \$6.7 million to the P-5 schools, which amounted to grants of approximately \$500 per child.¹⁹

To assess whether student achievement varies among the different types of public schools, I estimate the effect of the total number of years in which the student has continuously been enrolled, or had ever been enrolled, in the particular type of school.²⁰ Thus, I estimate the gap in test scores between

students in “regular” Milwaukee public schools and those enrolled in choice, citywide, and P-5 schools. I control for family background and student ability by including individual fixed effects, as described above.

Chart 2 shows the results for math scores.²¹ The differences in the top panel do not adjust for student ability and family background; those in the bottom panel do. Consider, first, the results that do not adjust for family background. These figures suggest that students in the citywide schools consistently score higher than students in the regular public schools, and the gap increases with the cumulative number of years the students have been enrolled in the citywide schools. This finding is consistent with much of the existing evidence on magnet schools (for example, Blank [1990], Crain, Heebner, and Si [1992], and Gamoran [1996]). In addition, the results indicate that although the students in the P-5 and choice schools have lower scores (than students in the regular public schools) in the first year, the rate of increase is (roughly) similar to that

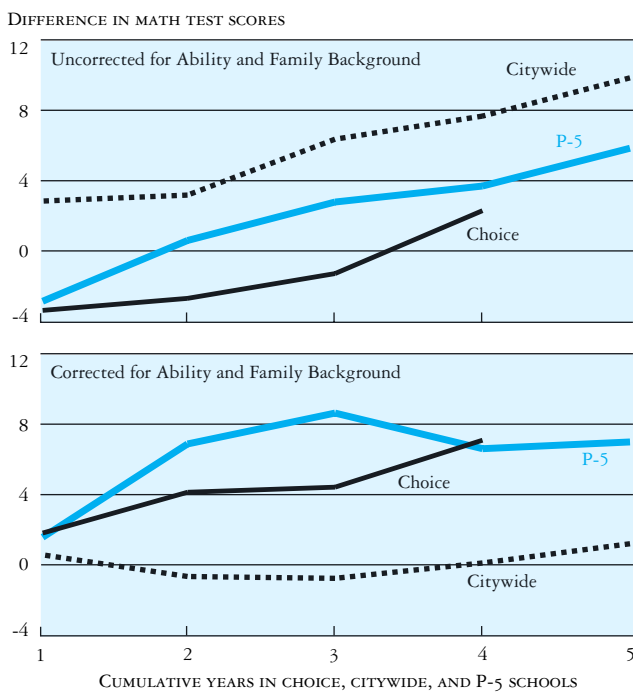
for students in the magnet schools.²²

Notice, however, the effect of controlling for student ability and family background using individual fixed effects, as shown in the bottom panel of Chart 2. Once student characteristics have been accounted for, the gap in math scores between the citywide students and regular public school students disappears. At the same time, the gap between those in the P-5 and choice schools becomes large and statistically meaningful. Significantly, there is no difference in the math achievement gains of students in the P-5 and choice schools.

Chart 3 presents the reading score results. Again, before controlling for student ability and background (with individual fixed effects), I find that students in citywide schools score substantially higher than students in the regular public schools and in the choice schools (top panel). Students in the P-5 schools make incremental yearly gains, although these gains are not statistically distinguishable from zero. The bottom panel again shows that once one adjusts for individual

Chart 2

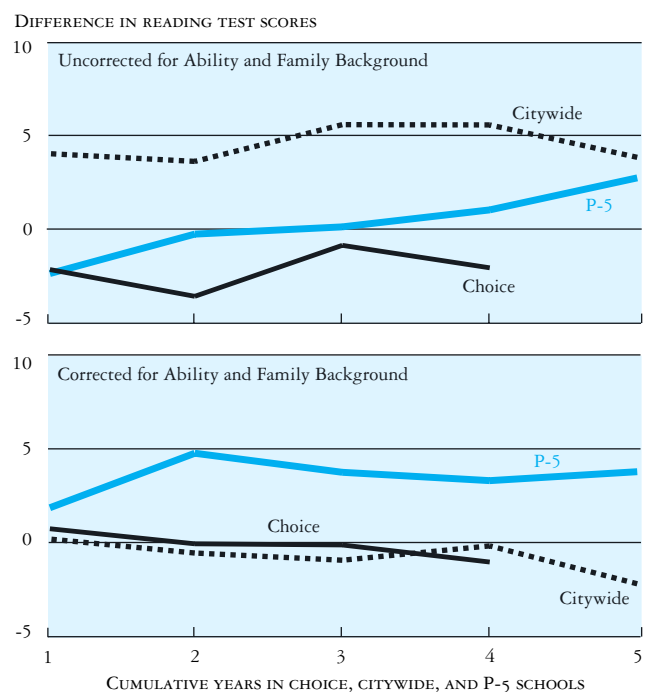
THE DIFFERENCE IN MATH TEST SCORES BETWEEN CHOICE, CITYWIDE, AND P-5 SCHOOLS, AND “REGULAR” MILWAUKEE PUBLIC SCHOOLS



Notes: The top panel does not control for individual fixed effects; the bottom panel does. A P-5 school participates in Milwaukee’s Preschool to Grade 5 Grant Program.

Chart 3

THE DIFFERENCE IN READING TEST SCORES BETWEEN CHOICE, CITYWIDE, AND P-5 SCHOOLS, AND “REGULAR” MILWAUKEE PUBLIC SCHOOLS



Notes: The top panel does not control for individual fixed effects; the bottom panel does. A P-5 school participates in Milwaukee’s Preschool to Grade 5 Grant Program.

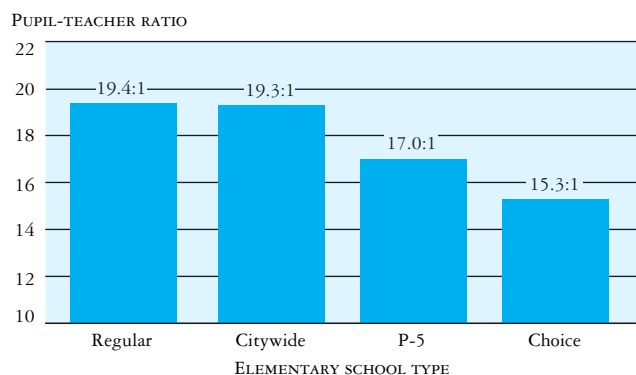
fixed effects, students in the citywide and choice schools are found not to have faster reading test score gains than students in the regular public schools.²³ In contrast, students in the P-5 schools have substantially faster gains in reading than those in the other public schools and choice schools.

Overall, these results suggest that the observed superiority of the citywide schools in Milwaukee can be attributed to the fact that they enroll higher achieving students.²⁴ The results also suggest that students in the P-5 schools have math score gains equal to those of students in the choice schools and reading score gains that are greater. After four years, the P-5 and choice test score advantage is about 0.37 of a standard deviation for math; the P-5 advantage in reading is about 0.16 of a standard deviation.²⁵ These gains are relatively large for education productions, and are comparable to the effects from the Tennessee class size experiment (Finn and Achilles 1990; Krueger 1997).

What might explain the fact that the P-5 and choice schools generally outperform the other public schools? While there are undoubtedly many factors that might explain this result, one relatively easily observed characteristic that they have in common is a small pupil-teacher ratio, which is often used as a proxy for class size.²⁶ Chart 4 shows the average pupil-teacher ratio by school type.²⁷ The average pupil-teacher ratio in the P-5 schools is 17.0 students per teacher; the average ratio in the choice schools is 15.3.

Chart 4

AVERAGE PUPIL-TEACHER RATIO BY ELEMENTARY SCHOOL TYPE



Notes: Ratios are enrollment-weighted. A P-5 school participates in Milwaukee's Preschool to Grade 5 Grant Program.

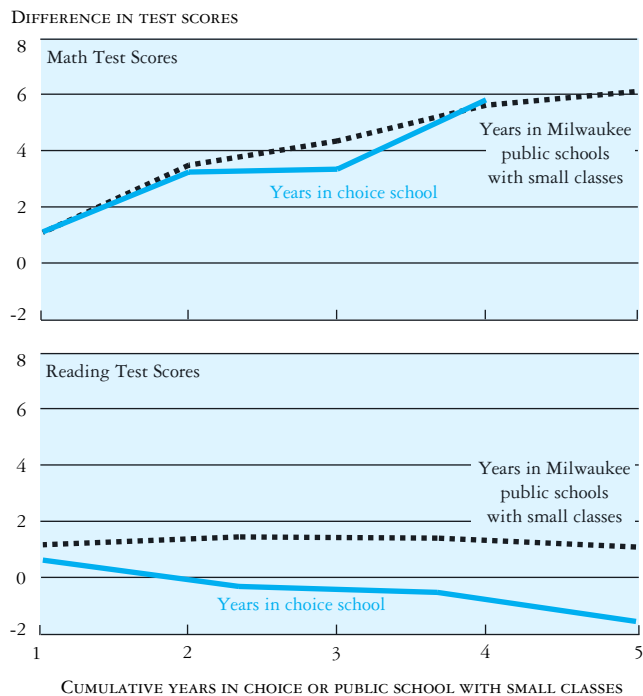
Both are significantly smaller than the pupil-teacher ratios in the regular and citywide public schools.

To gauge the extent to which small pupil-teacher ratios might explain the achievement effects of the choice program, I first estimate the effect of the choice schools on test scores relative to all Milwaukee public schools. Next, I estimate the achievement gains that accrue to students enrolled in public schools with low pupil-teacher ratios relative to those in public schools with higher pupil-teacher ratios.²⁸ This latter analysis uses only students enrolled in the Milwaukee public schools. I then compare the two sets of achievement gains. A finding that the gains among the public schools with low pupil-teacher ratios largely correspond to the gains in the choice schools provides *indirect* evidence that low pupil-teacher ratios (and perhaps small class sizes) may explain part of the observed private-public school achievement differentials (Chart 5).²⁹

The solid line in the top panel of Chart 5 shows the math test score growth of students in the choice schools relative to students in all Milwaukee public schools. These results essentially replicate those presented in the top panel of Chart 1. The dotted line shows the math test score progression of students in public schools with small pupil-teacher ratios relative to students in public schools with larger pupil-teacher ratios. The two lines almost entirely coincide. The results for reading are in the bottom panel. In this case, the two lines do not overlap to the same degree as those for math; however, none of the gaps is significantly different from zero, either.

These results indicate that lower pupil-teacher ratios (or class sizes) *may* explain the differential math gains by students in the choice schools (as well as the lack of gains in reading). They do not, however, explain why the P-5 schools appear to perform so well in reading. It is important to understand that this exercise does not *prove* that low pupil-teacher ratios explain either the public-private school or the P-5–regular school achievement difference. Rather, this exercise highlights the need for a much better understanding of why the choice schools in Milwaukee may, on average, be better (at least in teaching mathematics) than the average public school, and why the P-5 schools appear as strong as the choice schools and stronger than the

CAN SMALL CLASSES EXPLAIN THE CHOICE SCHOOL ACHIEVEMENT EFFECTS?



Notes: The chart depicts the difference in test scores between choice schools and all Milwaukee public schools, and between public schools with small pupil-teacher ratios and public schools with larger pupil-teacher ratios (controlling for ability and family background with individual fixed effects). Milwaukee public schools with small classes are those with a pupil-teacher ratio less than or equal to seventeen students per teacher.

other public schools. I have looked at pupil-teacher ratios because they are a readily available measure that partially defines the P-5 schools and because representatives from the choice schools I contacted emphasized their small class sizes. However, there are likely to be other equally compelling school-specific factors that may explain the differences. Moreover, it is critical that we understand these factors better if we are to improve education for America’s urban youth.

CONCLUSION

The results in this paper suggest that there are significant differences between the public schools in Milwaukee. In particular, students who attend a subset of schools distinguished by, among other characteristics, their small class sizes

and additional state funding have test score gains in math that keep pace with those in the private schools that participate in the Milwaukee Parental Choice Program. In addition, this subset of schools has significantly faster reading score gains than either the choice or the other public schools.

In order to evaluate these results fully, one must consider not only student achievement, but costs as well. Evidence that students performed just as well (or better) in the choice schools, but at lower cost, would indicate that private schools are more efficient. Unfortunately, I know of no definitive accounting of the cost differences between the two sectors for providing the same mix of services. Some researchers argue that private schools cost 50 to 60 percent less than public schools (for example, Hoxby [1998]). Coleman and Hoffer (1987) report that, among high schools, overall private school expenditures are 91 percent of public school expenditures. However, both “other, non-Catholic” and “high-performance” private schools spend more than public schools.³⁰ In addition, Levin (forthcoming) presents an extremely rough estimate of the costs in the Milwaukee public and choice schools. He concludes that the choice schools may have only slightly lower costs (for the same services). Therefore, particularly compared with the P-5 schools, the choice schools may not have an unambiguous efficiency advantage. Clearly, a careful comparison of the educational costs in public and private schools would make an invaluable contribution to the literature and the public policy discussion.

This analysis provides direct evidence that not all public schools are created equal. In addition, not all private schools are created equal. For example, while the overall results suggest that students in the choice schools have no faster gains in reading than do students in the (average) Milwaukee public school, Hispanic students in the choice program—90 percent of whom attend one private school—do make significant gains in reading.³¹ If we really want to “fix” our educational system, we need a better understanding of what makes a school successful, and we should not simply assume that market forces explain sectoral differences and are therefore the magic solution for public education.

APPENDIX

Table A1
ELEMENTARY AND MIDDLE SCHOOLS CLASSIFIED
AS P-5 AND CITYWIDE

P-5	Citywide
Auer*	Brown
Clarke*	Craig
Franklin*	Elm
Green Bay	Fratney
Holmes	Garfield Avenue
Hopkins*	Grant Avenue
Kagel*	Greenfield
Keefe*	Hawley Road
Kilbourne	Lincoln Center for the Arts
LaFollette*	Lloyd
Lee*	MacDowell
Martin Luther King, Jr.*	Meir Elementary School
Palmer	Milwaukee Education Center
Phillis Wheatley*	Milwaukee French Immersion
Pierce*	Milwaukee German Immersion
Riley	Milwaukee Spanish Immersion
Siefert*	Morgandale
Thirty-first Street (Westside)*	Morse
Thirty-seventh Street	Robinson
Twenty-seventh Street*	Roosevelt
Vieau	Sara Scott
	Starms Discovery
	Thirty-eighth Street
	Thurston Woods
	Tippecanoe
	Townsend Street
	Twenty-first Street
	Urban Waldorf

Note: A P-5 school participates in the Preschool to Grade 5 Grant Program.

*Denotes an original Project Rise School.

Table A2
ORDINARY LEAST SQUARES AND INDIVIDUAL FIXED-EFFECTS
ESTIMATES OF THE EFFECT OF CITYWIDE, P-5, AND CHOICE
SCHOOLS ON MATH AND READING TEST SCORES

	Dependent Variable			
	Math Scores		Reading Scores	
	Ordinary Least Squares	Fixed- Effects	Ordinary Least Squares	Fixed- Effects
Currently enrolled in citywide school	4.240 (1.105)	0.233 (1.054)	2.565 (1.000)	-1.033 (0.986)
Enrolled one year	-1.375 (1.094)	0.347 (1.308)	1.471 (0.996)	1.245 (1.226)
Enrolled two years	-1.037 (1.212)	-0.882 (1.411)	1.062 (1.091)	0.504 (1.336)
Enrolled three years	2.139 (1.401)	-0.986 (1.533)	3.018 (1.264)	0.121 (1.438)
Enrolled four years	3.448 (1.975)	-0.114 (1.878)	3.003 (1.765)	0.883 (1.755)
Enrolled five years	5.620 (3.142)	0.978 (2.683)	1.279 (2.679)	-1.134 (2.392)
Currently enrolled in P-5 school	2.446 (0.850)	1.810 (0.688)	3.529 (0.762)	0.439 (0.649)
Enrolled one year	-5.271 (0.741)	-0.234 (1.743)	-5.871 (0.670)	1.441 (1.623)
Enrolled two years	-1.821 (0.885)	5.067 (1.777)	-3.773 (0.798)	4.342 (1.653)
Enrolled three years	0.363 (1.071)	6.820 (1.893)	-3.388 (0.953)	3.328 (1.759)
Enrolled four years	1.271 (1.367)	4.799 (2.054)	-2.483 (1.159)	2.885 (1.914)
Enrolled five years	3.417 (2.014)	5.180 (2.372)	-0.779 (1.746)	3.361 (2.222)
Currently enrolled in choice school	0.338 (1.739)	-2.631 (1.391)	0.297 (1.547)	-1.558 (1.331)
Enrolled one year	-3.683 (1.656)	4.450 (1.762)	-2.428 (1.484)	2.321 (1.673)
Enrolled two years	-2.999 (1.844)	6.766 (1.839)	-3.853 (1.651)	1.519 (1.743)
Enrolled three years	-1.592 (2.193)	7.054 (2.045)	-1.139 (1.980)	1.458 (1.943)
Enrolled four years	1.980 (3.113)	9.721 (2.560)	-2.336 (2.797)	0.549 (2.421)

Memo:

Control for individual fixed effects?	No	Yes	No	Yes
R ²	0.057	0.819	0.039	0.795
Number of observations	10,186	10,186	10,224	10,224

Notes: Standard errors are in parentheses. All specifications include a constant and dummy variables indicating the grade level of the student when he or she took the test. The math score regressions include a dummy variable indicating if the test score was imputed. "Enrolled" is the total number of years the student has continuously been enrolled, or had ever been enrolled, in the particular type of school. A P-5 school participates in the Preschool to Grade 5 Grant Program.

APPENDIX (*Continued*)

Table A3
 INDIVIDUAL FIXED-EFFECTS ESTIMATES OF THE EFFECT
 OF CHOICE SCHOOLS AND PUBLIC SCHOOLS WITH SMALL PUPIL-
 TEACHER RATIOS ON MATH AND READING TEST SCORES

	Sample of Choice and Public Schools		Sample of Only Public Schools	
	Dependent Variable			
	Math Scores (1)	Reading Scores (2)	Math Scores (3)	Reading Scores (4)
Currently enrolled in school with small pupil-teacher ratio			-0.232 (1.111)	-2.383 (1.022)
Enrolled one year			1.333 (1.360)	3.542 (1.254)
Enrolled two years			3.725 (1.518)	3.821 (1.401)
Enrolled three years			4.585 (1.727)	3.778 (1.593)
Enrolled four years			5.851 (1.985)	3.463 (1.829)
Enrolled five years			6.332 (2.294)	4.247 (2.101)
Currently enrolled in choice school	-3.459 (1.365)	-2.312 (1.297)		
Enrolled one year	4.584 (1.734)	2.926 (1.651)		
Enrolled two years	6.707 (1.813)	1.992 (1.713)		
Enrolled three years	6.810 (2.012)	1.781 (1.903)		
Enrolled four years	9.269 (2.526)	0.749 (2.376)		
Memo:				
R ²	0.816	0.795	0.819	0.803
Number of observations	10,186	10,224	7,171	7,241

Notes: Standard errors are in parentheses. All specifications also include a constant and dummy variables indicating the grade level of the student when he or she took the test, and individual fixed effects. The math score regressions include a dummy variable indicating if the test score was imputed. The regressions in columns (1) and (2) compare the choice schools with all Milwaukee public schools; those in columns (3) and (4) include only the Milwaukee public schools. "Enrolled" is the total number of years the student has continuously been enrolled, or had ever been enrolled, in the particular type of school.

ENDNOTES

The author thanks Alan Krueger and Michele McLaughlin for useful conversations and Howard Fuller for helping her to classify (and understand) the Milwaukee public schools. Michele McLaughlin also provided expert research assistance. Any errors are the author's.

1. For excellent descriptions of the program, see Witte, Thorn, Pritchard, and Claibourn (1994) and Witte, Sterr, and Thorn (1995).
2. As a result, the schools participating in the voucher program are not representative of the typical private school, since only 21 percent of private schools are nonsectarian (U.S. Department of Education 1996). However, until the constitutionality of whether religious schools can participate in voucher programs has been decided, the experience in Milwaukee will be relevant for other cities considering such reforms.
3. I obtained this information by calling the five schools enrolling the largest proportion of choice students. Combined, these schools enroll over 95 percent of the choice students.
4. Originally, the private schools in the choice program were only allowed to admit up to 49 percent of their students as part of the program; this level was raised to 65 percent in 1994. In addition, the number of students who could participate in the choice program was originally limited to 1 percent of the Milwaukee public school enrollment in the first four years but was increased to 1.5 percent in 1994. Given the total enrollment in the Milwaukee public schools, there could be a maximum of only about 1,000 students in the program at any one time.
5. The term control group is generally reserved for randomized experiments, while comparison groups are developed from survey or administrative data.
6. In most other settings, the comparison would show that students in private schools outperform those in public schools.
7. In principle, if one had measures of all the characteristics in which students in the choice schools and students in the public schools differed, one could simply control for these and generate the true effect of the program. The problem, however, is that one is never sure that every characteristic has been controlled for, and indeed we rarely have measures of all (relevant) aspects of the students and their parents. With application lotteries, one does not need these measures.
8. One must control for the application lotteries because applicants to some schools were more likely to be selected than applicants to other schools.
9. There are several places where there could be slippage between the actual lotteries and the imputations. For example, children with siblings who are already enrolled in a choice school are exempted from the lottery, children can apply to more than one school at a time, and the Greene, Peterson, and Du (1997) imputation assumes that a child's race completely determines the school to which he or she applies. In addition, Witte (1997) expresses concern that the choice schools may have abused the permitted exclusions in order to have more control over which students they enrolled.
10. Another way to think about this diagram is that it represents the test score trajectories for both students in the absence of the choice program.
11. The test scores used in this paper are the normal curve equivalent scores of the Iowa Tests of Basic Skills. See Rouse (forthcoming) for more information about the sample.
12. See Rouse (1997) or Rouse (forthcoming) for an elaboration of these points.
13. The fact that the individual fixed-effects strategy can accommodate students missing prior test scores appears to explain a significant portion of the difference in our results.
14. Beginning in 1993, there was no "total math score" (from the Iowa Tests of Basic Skills) for a substantial percentage of students in the Milwaukee public schools. Therefore, I predict (or impute) the total score from the subset of students in the Milwaukee public schools who took the entire battery of math tests (see Rouse [forthcoming] for more details).
15. Others have argued that the observed private school effect is due to the selection process that leads higher achieving students to attend private schools. That is, they argue that the researchers have not controlled for all of the differences between the students in the private schools and the comparison group of students in the public schools. (See, for example, Cain and Goldberger [1983], Cookson [1993], Goldberger and Cain [1982], Murnane [1984], and Witte [1992]).
16. Ideally, I would also disaggregate the achievement gains by the individual choice schools. However, the state of Wisconsin has asked that such an analysis not be undertaken in order to preserve the confidentiality of the choice students.
17. This is my calculation, based on the Common Core of Data for 1991-92.
18. See Table A1 in the appendix for a list of the schools categorized as P-5 and citywide.

ENDNOTES (*Continued*)

19. This is my calculation, based on the Common Core of Data for 1991-92.

20. I estimate the effect of being enrolled in—rather than being selected to attend—the different types of schools because the results in Rouse (forthcoming) suggest that the analyses yield similar results. In addition, estimating the effect of being selected to attend the different types of public schools (and estimating the effect of “years since application”) does not make as much sense.

21. The underlying coefficient estimates and standard errors for Charts 2 and 3 are in Table A2 in the appendix.

22. The gap between students in the P-5 and regular public schools becomes statistically significant in the third year. The gap for the choice schools is not statistically significant.

23. These results differ from those reported by Archbald (1995), who found that students in the Milwaukee magnet schools scored higher on math and reading tests than those enrolled in the attendance area schools.

24. Not all citywide schools perform equally. In particular, when these schools are divided into “gifted,” “language immersion,” “special program” (such as Waldorf, Montessori, or Global Learning), and “other”—and individual fixed effects are included—the students in the language immersion schools have substantially faster gains in reading than students in all other types of schools, and students in the gifted schools have significantly slower gains in mathematics than students in the regular schools. Because in some years the number of students in some of these school categories is small, these results should be regarded as tentative.

25. I used the within-sample standard deviation of 19 for this calculation. Nationally, the standard deviation for normal curve equivalent scores is 21.

26. Although highly correlated, the pupil-teacher ratio does not always directly correspond to the average class size. Rather, the two measures diverge as intraschool variation in class size increases due, for example, to special and compensatory education (Boozar and Rouse 1997). To illustrate, the average pupil-teacher ratio in the choice schools is 15.3 students per teacher; however, the schools’ average class size is 23.6 students. Unfortunately, data on average class size for the Milwaukee public schools were not readily available.

27. The estimates of the pupil-teacher ratios for the choice schools are based on the schools I contacted. I estimated the pupil-teacher ratios for the public schools using the Common Core of Data for 1991-92.

28. Schools with low pupil-teacher ratios have ratios less than or equal to seventeen to one. I chose seventeen because it is the maximum pupil-teacher ratio in the choice schools I contacted. According to this criterion, 43 percent of all Milwaukee public schools and 52 percent of the P-5 schools are considered to have low pupil-teacher ratios.

29. See Table A3 in the appendix for the estimated coefficients and standard errors.

30. Other, non-Catholic private schools spend 38 percent more than public schools, while high-performance private schools spend 131 percent more than public schools. Expenditures in Catholic private schools, however, are lower than those in public schools. The fact that Catholic school costs differ from those in other types of private schools may reflect lower teacher salaries and greater in-kind subsidies (including facilities) from the Catholic church.

31. These results are not reported here but are available from the author on request.

REFERENCES

- Archbald, Doug.* 1995. "A Longitudinal Cohort Analysis of Achievement Among Elementary-Magnet Students, Neighborhood-School Students, and Transfer Students." *JOURNAL OF RESEARCH AND DEVELOPMENT IN EDUCATION* 28, no. 3 (spring): 161-8.
- Blank, Rolf K.* 1990. "Analyzing Educational Effects of Magnet Schools Using Local District Data." *SOCIOLOGICAL PRACTICE REVIEW* 1, no. 1: 40-51.
- Boozer, Michael A., and Cecilia Elena Rouse.* 1997. "Intra-School Variation in Class Size: Patterns and Implications." Revision of NBER Working Paper no. 5144 (June 1995).
- Bryk, Anthony S., Valerie E. Lee, and Peter B. Holland.* 1993. *CATHOLIC SCHOOLS AND THE COMMON GOOD*. Cambridge: Harvard University Press.
- Cain, Glen G., and Arthur S. Goldberger.* 1983. "Public and Private Schools Revisited." *SOCIOLOGY OF EDUCATION* 56 (October): 208-18.
- Clancy, Dan, Charlie Toulmin, and Merry Bukolt.* 1995. "Wisconsin Public School Finance Programs, 1993-94." In Steven D. Gold, David M. Smith, and Stephen B. Lawton, eds., *PUBLIC SCHOOL FINANCE PROGRAMS OF THE UNITED STATES AND CANADA 1993-94*. New York: American Education Finance Association.
- Coleman, James S., and Thomas Hoffer.* 1987. *PUBLIC AND PRIVATE HIGH SCHOOLS: THE IMPACT OF COMMUNITIES*. New York: Basic Books.
- Coleman, James S., Thomas Hoffer, and Sally Kilgore.* 1982a. *HIGH SCHOOL ACHIEVEMENT: PUBLIC, CATHOLIC, AND PRIVATE SCHOOLS COMPARED*. New York: Basic Books.
- . 1982b. "Cognitive Outcomes in Public and Private Schools." *SOCIOLOGY OF EDUCATION* 55, no. 2-3 (April-July): 65-76.
- Cookson, Peter W.* 1993. "Assessing Private School Effects: Implications for School Choice." In Edith Rasell and Richard Rothstein, eds., *SCHOOL CHOICE: EXAMINING THE EVIDENCE*. Washington, D.C.: Economic Policy Institute.
- Crain, Robert L., Amy L. Heebner, and Yiu-Pong Si.* 1992. "The Effectiveness of New York City's Career Magnet Schools: An Evaluation of Ninth Grade Performance Using an Experimental Design." National Center for Research in Vocational Education working paper.
- Evans, William N., and Robert M. Schwab.* 1995. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *QUARTERLY JOURNAL OF ECONOMICS* 110 (November): 941-74.
- Finn, Jeremy D., and Charles M. Achilles.* 1990. "Answers and Questions About Class Size: A Statewide Experiment." *AMERICAN EDUCATIONAL RESEARCH JOURNAL* 27, no. 3 (fall): 557-77.
- Gamoran, Adam.* 1996. "Student Achievement in Public Magnet, Public Comprehensive, and Private City High Schools." *EDUCATIONAL EVALUATION AND POLICY ANALYSIS* 18, no. 1 (spring): 1-18.
- Goldberger, Arthur S., and Glen G. Cain.* 1982. "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer and Kilgore Report." *SOCIOLOGY OF EDUCATION* 55, no. 2-3 (April-July): 103-22.
- Greene, Jay P., Paul E. Peterson, and Jiangtao Du.* 1997. "The Effectiveness of School Choice: The Milwaukee Experiment." Harvard University Education Policy and Governance Occasional Paper no. 97-1, March.
- Hoxby, Caroline M.* 1998. "What Do America's 'Traditional' Forms of School Choice Teach Us about School Choice Reforms?" Federal Reserve Bank of New York *ECONOMIC POLICY REVIEW* 4, no. 1.
- Krueger, Alan B.* 1997. "Experimental Estimates of Education Production Functions." Princeton University Industrial Relations Section Working Paper no. 379, May.
- Levin, Henry M.* Forthcoming. "Educational Vouchers: Effectiveness, Choice, and Costs." *JOURNAL OF POLICY ANALYSIS AND MANAGEMENT*.
- Milwaukee Public Schools.* 1997. *DIRECTIONS: YOUR SCHOOL SELECTION GUIDE FOR THE 1997-98 SCHOOL YEAR*.
- Molnar, Alex.* 1996. *GIVING KIDS THE BUSINESS: THE COMMERCIALIZATION OF AMERICA'S SCHOOLS*. Boulder, Colo.: Westview Press.
- Murnane, Richard J.* 1984. "A Review Essay—Comparisons of Public and Private Schools: Lessons from the Uproar." *JOURNAL OF HUMAN RESOURCES* 19: 263-77.
- Neal, Derek.* 1997. "The Effect of Catholic Secondary Schooling on Educational Achievement." *JOURNAL OF LABOR ECONOMICS* 15, no. 1, part 1 (January): 98-123.
- Rouse, Cecilia Elena.* 1997. "Lessons from the Milwaukee Parental Choice Program." *POLICY OPTIONS* 18, no. 6 (July-August): 43-6.
- . Forthcoming. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *QUARTERLY JOURNAL OF ECONOMICS*.

REFERENCES (*Continued*)

Sander, William. 1996. "Catholic Grade Schools and Academic Achievement." *JOURNAL OF HUMAN RESOURCES* 31 (summer): 540-8.

U.S. Department of Education. National Center for Education Statistics. 1996. *DIGEST OF EDUCATION STATISTICS, 1996.* NCES 96-133, by Thomas D. Snyder, Charlene M. Hoffman, and Claire M. Geddes. Washington, D.C.

Witte, John F. 1992. "Private School vs. Public School Achievement: Are There Findings That Should Affect the Educational Choice Debate?" *ECONOMICS OF EDUCATION REVIEW* 11 (December): 371-94.

———. 1997. "Achievement Effects of the Milwaukee Voucher Program." Unpublished paper, University of Wisconsin, January.

Witte, John F., Troy D. Sterr, and Christopher A. Thorn. 1995. "Fifth-Year Report: Milwaukee Parental Choice Program." Unpublished paper, University of Wisconsin, December.

Witte, John F., Christopher A. Thorn, Kim M. Pritchard, and Michele Claibourn. 1994. "Fourth-Year Report: Milwaukee Parental Choice Program." Unpublished paper, University of Wisconsin, December.

Summary of Floor Discussion

David Brauer

The second session's discussion focused on the evaluation of school choice programs and the relationship between a school district's perceived quality of education and parental selection of neighborhoods. In particular, the participants discussed several factors—such as family background, suburban flight, and per pupil spending—that may have contributed to the results in the session's two papers and in other related research.

Derek Neal began by questioning an earlier study by John Witte on the Milwaukee Parental Choice Program, which Cecilia Rouse had cited in her paper. Witte's findings suggested that once a student's prior test scores were taken into account, there was no discernible difference in performance between the Milwaukee public schools and the private schools in the Milwaukee choice program. Neal said that because of errors in measuring test scores, Witte's results could have been biased against finding that Milwaukee's choice schools had a positive effect. Rouse replied that Witte had indeed not attempted to address this measurement issue.

Howard Chernick followed by asking whether the flight of relatively high-income families from Milwaukee to the suburbs might have skewed Rouse's evaluation of the Milwaukee program by removing above-average students from the system. Rouse responded that had such flight occurred after the program's adoption, it would have made the choice schools appear to be doing relatively better and the regular Milwaukee public schools seem to

be doing relatively worse. She pointed out, however, that to the best of her knowledge the flight phenomenon largely predated Milwaukee's adoption of a school voucher program.

Next, Jean McConnell sought clarification of the effect of family background on Rouse's findings. Rouse stressed that family background is a real, important effect that analysts must take into account when evaluating choice programs. When one does *not* control for family background, she said, the choice schools appear to be performing poorly—mostly because students eligible for the program come from less affluent families; when one does control for it, the choice schools perform better. Eric Hanushek then pointed out that in Milwaukee the choice schools spend only half as much per pupil as the public schools. Rouse noted that the figures cited by Hanushek significantly understate costs for the choice schools because they do not include federal subsidies and other income sources, omissions that make it difficult to determine which type of school actually spends more per pupil.

The discussion turned to Caroline Hoxby's paper. Ann Davis asked whether Hoxby's finding that families tend to move into districts with good schools can be clearly separated from the tendency of families to choose to live in neighborhoods where other families have similar social, financial, and racial characteristics. Hoxby said that this is an important consideration, but stressed that her research enables one to separate these effects because the ability to

choose a neighborhood based on the characteristics of its residents is always present, while the ability to choose a public school district within a metropolitan area may not always be. Parental choice of neighborhoods, she noted, will have more impact on the resources available to their children's schools in metropolitan areas with many small school districts (such as Boston) than in metropolitan areas with little or no choice of school districts (such as Miami). Hoxby also stressed that because many neighborhoods and schools are already highly segregated, the adoption of parental choice programs would not necessarily lead to further segregation.

Bill Andrews then asked whether Hoxby's earlier evaluations of charter schools had controlled for differences in the nature of such programs. Hoxby acknowledged that this is a very important question—adding that the exact nature of charter schools varies considerably across states in terms of degree of autonomy, financial independence, and continued exposure to regulation. While it is too soon to draw firm conclusions, she said, these differences are likely

to be quite important in the evaluation of charter schools, and some arrangements will almost certainly prove more effective than others.

Finally, Ronnie Lowenstein noted that per pupil spending tends to be lower in school districts where voters must approve the annual school budget—a pattern that could in part reflect the influence of elderly voters, whose children are no longer in the school system. Lowenstein wondered whether Hoxby's results were driven by the fact that voters in large districts are less likely to have control over school budgets than voters in small districts. Hoxby agreed that it is very important to understand how financial decisions are made. She pointed out, however, that although per pupil spending tends to be lower in small school districts, student achievement there appears to be higher than in large school districts. Nonetheless, Hoxby expressed concern over recent findings that districts with large numbers of elderly voters—particularly where these voters are not of the same race as school-age children—tend to support very low levels of public school spending.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

What Have We Learned about the Benefits of Private Schooling?

Derek Neal

In 1980, the U.S. Education Department, working with the National Opinion Research Center, began a panel study of high school students known as the “High School and Beyond Study.” The first wave of the study collected achievement test scores for approximately 50,000 high school students who were in either their sophomore or their senior year of high school. The survey also compiled detailed characteristics of schools and, for more than half the students involved, detailed follow-up surveys concerning school, work, and other activities.

The survey design yielded a sample that included a disproportionate share of private schools. When James Coleman of the University of Chicago took on the task of evaluating the first wave of data, he decided to exploit this unique aspect of the survey. In 1981, Coleman and two of his colleagues, Sally Kilgore and Thomas Hoffer, presented a report to the

National Center for Education Statistics entitled *Public and Private High Schools*, which concluded that the selection of superior students into private schools cannot explain the higher levels of achievement in private schools. Therefore, the authors argued that Catholic and other private schools are, as a rule, more effective institutions of learning than public schools.

This report and subsequent publications by Coleman and his associates ignited a heated and often acrimonious debate among social scientists concerning the relative educational performance of public and private schools. For example, Coleman, Kilgore, and Hoffer (1981) found that, in a population of students from similar backgrounds, private school students exhibit higher achievement and attainment. Critics, however, claimed that this result simply reflected inadequate controls for the individual traits and family background characteristics that foster academic success. Put simply, even if one knows a considerable amount about a student’s background and academic aptitude, the fact that her parents are willing to spend

Derek Neal is an associate professor of economics at the University of Chicago, a faculty research fellow at the National Bureau of Economic Research, and a faculty affiliate of the Joint Center for Poverty Research.

their time and resources to send her to a private school may provide additional information about the student's academic ability and family environment.

Because the 1981 report made controversial claims, it served as a catalyst for research on the relative performance of public and private schools. This paper attempts to summarize this research and also to assess what we have learned since 1981. Although many questions remain unanswered, one result seems clear. Black and Hispanic students in large cities often have the most to gain from private schooling, in particular, Catholic schooling. Further, the poor quality of many inner-city schools appears to drive this result.

The balance of the paper reviews results concerning private schooling's effect on academic achievement and attainment. I pay particular attention to the literature on Catholic school effects because Catholic schools constitute a large and relatively homogeneous set of private schools. I then discuss the implications for the ongoing debate over vouchers. I conclude with some thoughts about future research.

AT LEAST ONE CONSISTENT PATTERN

In the literature on the effects of private schooling, many results appear quite fragile. Estimates of the achievement gains associated with private schooling often vary considerably across studies that employ the same data sources. One result, however, remains constant across a number of studies that vary with respect to data sources and methodology. For many students, Catholic secondary schooling raises graduation rates.

The table on this page presents a summary of results from three studies. Evans and Schwab (1995) and Sander (1997) use the 1986 follow-up survey to the "High School and Beyond Study."¹ Neal (1997) uses data from the National Longitudinal Survey of Youth (NLSY). For each study, the table provides results from single equation models that explain high school graduation as a function of numerous family background characteristics and, in some instances, prior measures of achievement. Neal reports the effect of attending a Catholic secondary school on the probability of graduation. The other two studies report the effect of Catholic schooling on graduation rates, given that

students stay in Catholic school through the spring semester of tenth grade.

Evans and Schwab find that Catholic schooling increases graduation rates. According to their results, Catholic schooling raises graduation rates by 14 percentage points for whites and 13 percentage points for blacks. Further, Evans and Schwab find even larger gains from Catholic schooling when they restrict their attention to students in urban areas.

Neal's approach is slightly different because it employs sample definitions that involve both race and community type. The results suggest that urban minorities attending Catholic secondary schools experience a 26 percentage point increase in the probability of graduating from high school. The corresponding figure for urban whites is 10 percentage points. Neal's results for nonurban students, which are not reported here, indicate smaller and statistically insignificant effects on graduation rates regardless of race.

Sander examines graduation rates for rural students in the United States. He finds that, in towns outside major metropolitan areas, Catholic schooling has small and statistically insignificant effects on graduation rates.

These three studies all attempt to correct their single equation model estimates for selection bias. Single equation estimates may be misleading because unmeasured traits that make students more likely to attend Catholic schools may also make them more likely to graduate. I do not present the corrected estimates here because none of these studies finds significant evidence of selection bias. In fact, most studies report weak evidence

CHANGES IN THE PROBABILITY OF HIGH SCHOOL GRADUATION

Source	Sample	Graduation Rate (Percent)	Percentage Point Increase in Graduation Rate (Catholic Schooling Effect)
Evans and Schwab	White	83	14
	Black	80	13
	Urban	77	17
	Suburban	86	11
Neal	Urban: black and Hispanic	64	26
	Urban: white	76	10
Sander	Rural	82	6 ^a

^a Figure is not statistically significant.

of negative selection into Catholic schools, and none provides strong evidence that the unmeasured traits of Catholic school students make them more likely to succeed in school than observationally similar public school students.

I wish to focus on how the “effect” of Catholic schooling differs across demographic groups. There is clear evidence that the benefits of Catholic schooling vary with location. Both Evans and Schwab and Neal report that the benefits of Catholic schooling are greatest for students who live in heavily populated areas, while Sander finds that Catholic schooling has no effect on graduation rates in rural areas.

Location is not the whole story, however. Neal’s results for urban minorities are particularly striking. A 26 percentage point increase in the probability of graduating from high school is an enormous effect. Further, Neal argues that a likely explanation for the concentration of Catholic school benefits in urban minority communities lies in differences between public schools, not in differences between public and Catholic schools. Neal estimates predicted graduation rates for public school students as a function of family background and community type. He constructs these estimates separately for whites and nonwhites and finds that in counties with fewer than a half million people, whites and nonwhites from similar backgrounds graduate from public schools at similar rates.

Consistent with Evans and Schwab’s finding of large Catholic school effects in urban areas, the NLSY data reveal lower graduation rates for students of all races in large cities. However, the decline is much more dramatic for nonwhites. In cities, whites and nonwhites graduate at very different rates, and these differences cannot be accounted for by differences in family background. In short, the graduation rate of minorities in urban public schools is quite low when compared with the graduation rates of either urban whites or minorities who live in nonurban areas.

Coleman and Hoffer (1987) examine dropout rates between the tenth and twelfth grades. Holding constant observed background characteristics among white students, they report an 11 percentage point gap between the dropout rates for Catholic school and public school students. For

minorities and for students from disadvantaged backgrounds, the gap is between 1 and 6 percentage points larger.²

Evans and Schwab also examine rates of college attendance. They do not report separate estimates of Catholic school effects for different populations, but they do report that, on average, Catholic schooling raises college entrance rates by about 14 percentage points. Neal examines college graduation rates and again finds evidence of large Catholic school effects among urban minorities. Neal’s estimates suggest that Catholic schooling raises college graduation rates for urban minorities from 11 to 27 percent. Further, when Neal considers only high school graduates, the corresponding increase is from 16 to 30 percent. Neal reports significant, but slightly smaller, effects for urban whites.

Numerous studies report that Catholic schooling enhances educational attainment. In general, estimates of these Catholic school effects are always larger in samples restricted to urban residents and in most cases larger in samples restricted to minority students.³ Further, Neal reports that minorities in large cities appear to benefit most from Catholic schools because they face poor public school alternatives.

Because the pecuniary returns of education have increased dramatically over the past two decades, the gains in educational attainment imply significant gains in earnings. Neal reports that the gains in attainment he finds may raise adult wages among urban minorities by as much as 8 percent.

MIXED RESULTS ON ACHIEVEMENT

In their original paper, Coleman, Kilgore, and Hoffer (1981) reported that in reading and vocabulary, Catholic school sophomores are roughly two grade equivalents ahead of their public school counterparts, and in math, slightly more than two grade equivalents ahead. In addition, the authors found that roughly 60 percent of these achievement differentials reflected differences in family background and therefore concluded that Catholic schooling raises achievement by roughly one grade level. Further, the authors claimed that Catholic schooling effects are even larger for minority students and students from economically disadvantaged backgrounds.

Numerous scholars from different disciplines conducted replication studies that challenged the robustness of Coleman, Kilgore, and Hoffer's 1981 results, but I will not explore the details of this debate for two reasons.⁴ First, a proper summary would necessarily be long and tedious. Second, work with the 1982 follow-up data settled many of the points raised in the original debate. In 1982, the original 1980 sophomore cohort took another battery of achievement tests. Researchers were then able to estimate achievement models using prior achievement measures as a control. Coleman and Hoffer (1987), Willms (1985), and Alexander and Pallas (1985) all analyzed the achievement data from the follow-up study, and all three reported similar results. In verbal skills, mathematics, and writing, Catholic school students scored about .1 standard deviation higher than students in public schools with comparable family backgrounds and sophomore achievement. In science and civics, the effects of Catholic schooling on achievement did not appear to be statistically significant.

The debate concerning these results was primarily rhetorical. Both Willms and Alexander and Pallas (1983) claimed that .1 standard deviation represents a trivial gain. Coleman and Hoffer (1987) noted that given the scores of seniors in the 1980 survey, this gain represents approximately one grade equivalent. Thus, if the gains from Catholic schooling between the eighth and tenth grades are the same as the gains between tenth and twelfth, attending four years of Catholic high school generates a .2 standard deviation increase in achievement. In terms of public school grade equivalents, Catholic schools would, in some subjects, offer six years of achievement for four years of attendance.

Coleman and Hoffer (1987) also found that the effects of Catholic schooling on achievement growth are greatest for minority students and students from economically and socially disadvantaged backgrounds. Given the standard errors reported by the authors, however, these differences are not statistically significant in many instances.

Although the analyses of achievement in the first follow-up survey provided controls for prior achievement, the results may still be contaminated by selection bias. Holding current achievement constant, students who are highly motivated may still be more likely to attend private

schools. Coleman and Hoffer (1987) did perform tests for selection bias using the follow-up data and found little evidence that selection bias contaminated their results.

More recently, researchers have begun analyzing data from the 1988 National Educational Longitudinal Study. This panel study began with a cohort of students who were finishing eighth grade in the spring of 1988. Taken together, the 1988 survey and subsequent follow-up surveys provide achievement test scores for eighth, tenth, and twelfth graders. Figlio and Stone (1997) conducted an analysis of these achievement data. Given their strategy of correcting estimates for selection bias, they find that private schools with a religious affiliation do not enhance achievement in the population as a whole or within most subgroups. However, the authors do report large achievement gains for blacks and Hispanics who attend private religious schools, and they report the largest gains for blacks and Hispanics who live in large, central cities.⁵

A 1990 case study by RAND supports the claim that minority youth in large cities benefit from Catholic schooling. Hill, Foster, and Gendler (1990) compare regular public schools, magnet schools, and Catholic schools in inner-city neighborhoods in New York City. They also gathered data from some inner-city schools in Washington, D.C. The authors focused their data collection on eight New York City schools that all contain substantial numbers of minority students and students from economically disadvantaged families. The study devotes particular attention to students who attended Catholic schools through a privately funded scholarship program. According to the authors, "most scholarship recipients are black or Puerto Rican.... They tend to come from single-parent welfare homes and have poor scholastic records."

Although many scholarship students entered Catholic school performing below grade level, 82 percent graduated. This figure compares with 55 percent in the regular public schools and 66 percent in the representative magnet school. Further, 85 percent of the scholarship students took the Scholastic Aptitude Test; on average, they scored almost as well as the students who paid to attend the Catholic schools. They also scored an average of almost

90 points higher than the 50 percent of magnet school students who took the test and 160 points higher than the 33 percent who took the test in regular public schools. Obviously, these findings only provide information about a small set of schools in one city. However, they are part of a notable pattern of results in the literature.

ADDITIONAL DATA: VOUCHER PROGRAMS

So far, I have largely restricted my attention to studies comparing Catholic and public schools. This perspective reflects the fact that a large fraction of private secondary schools are Catholic schools and that the balance of the private secondary school market is quite heterogeneous. No other relatively homogeneous group of private schools is well represented in data sets that provide student background characteristics as well as individual achievement and attainment data. Further, samples of minority students in secular private schools are usually quite small.

However, in recent years a set of studies concerning achievement in private elementary schools and dealing with a large sample of minority students has received a great deal of attention. In 1990, the city of Milwaukee provided a limited number of vouchers for low-income families. These vouchers, worth roughly \$2,500 each, could be used at private secular elementary schools. The data from the follow-up studies contribute interesting information to the debates over the relative effectiveness of private versus public schools. Because the program did not provide vouchers for every family that applied to the program, the data cover families that wanted to participate but were not permitted to do so. Thus, the data provide a natural comparison group for the students who attended private schools under the program.

Unfortunately, different researchers have drawn different conclusions from analyses of the Milwaukee data. I will not go further into this debate here because another paper in this volume (Rouse 1998) addresses the issues at length. Nonetheless, a recent paper by Rouse (forthcoming) does provide credible evidence that access to private education increased the math scores of program participants, although Rouse finds no evidence of positive effects on reading achievement.

POLICY IMPLICATIONS

Any regular C-SPAN viewer knows that scholars and policy-makers often talk past one another and that on any given topic the conventional wisdom among politicians may not coincide with the opinions of the majority of scholars who work on the topic. However, when I look at the academic literature on the benefits of private schooling, I see themes that are also common in newspaper and magazine articles concerning proposals for school reform.

The most compelling evidence for positive private school effects comes from analyses of minority youth in cities. Further, if for no other reason than data availability, this is particularly true with respect to Catholic school effects. Given this result, it is interesting to note that many privately funded voucher programs and most proposals for publicly funded vouchers target minority youth in large cities and, in many instances, minority youth in large cities with a significant number of Catholic schools.

A recent issue of *Time* magazine profiled a privately funded voucher program in Philadelphia. Last year, John Cardinal O'Connor touched off considerable debate by offering to take the lowest performing 5 percent of New York's public school students out of overcrowded public schools and place them in Catholic schools. In exchange, the Cardinal asked the city to provide \$2,500 per student. In 1996, Cleveland began the first state-funded voucher plan that included religious schools.

Given the recent flurry of voucher proposals targeted toward inner-city youth, it is interesting to ask whether or not the existing evidence supports the hypothesis that voucher plans will be successful. I believe two words of caution are in order.

First, none of the studies discussed above fully deals with the fact that some students may be better suited for Catholic schools than others. It is hard to find evidence that urban Catholic school students are simply better students than their public school counterparts on some unobserved dimension. However, existing Catholic school students may be the students who have the most to gain from Catholic schooling. We may be safe in concluding that Catholic schools provide real benefits for their current students. Much harder to ascertain is

how many other students would benefit from Catholic schooling if given the opportunity. Would students from Muslim families benefit from Catholic schooling? Given available data, we cannot answer this question. At best, we may expect significant benefits from Catholic schooling for students who are quite similar to the existing population of Catholic school students.

However, even if we consider a voucher program aimed at inner-city neighborhoods where Catholic or other private schools already succeed, we cannot confidently expect positive outcomes for program participants if the program is large in scale. For the outside observer, it is hard to know exactly what makes some schools succeed while others fail. Large school voucher programs would likely mean the expansion of many existing private schools and the entry of many new private schools. How would this expansion and entry affect the quality of private schools or the quality of remaining public schools? I do not know, and available data shed little light on this question.

Nonetheless, I see no reason to be wary of small-scale voucher plans that target disadvantaged students in large cities. Small-scale plans should not affect the current function of either private or public schools. Moreover, by targeting vouchers toward economically disadvantaged students in cities, we would aid students who currently receive poor service from public schools.

FUTURE RESEARCH AND RELATED POLICY CONCERNS

I have argued that some students benefit more from private schooling than others simply because the public schools available to them are worse than those available to others. The notion of “available public schools” is problematic, however, because families choose where they live and thus choose the schools that are available.

Although existing research tells us little about how families make the joint decision of where to live and where to send their children to school, a recent paper by Nechyba (1997) points to the potential payoffs of such research. Nechyba constructs a simulation model that explores what might happen in terms of school choices and residential location choices under a full-scale voucher system. His most interesting result is that an important link between school choice and residential location exists. In his simulations, a voucher program may reduce residential segregation by income class. An elastic supply of private schools makes it possible for people to uncouple school choice and place of residence. Families can live near their jobs and let good schools come to them. Nechyba’s paper raises the possibility that a broad-based voucher program might also serve as an urban renewal program. How many commuters would decide to live in the cities where they work if they could use vouchers to choose from a menu of private schools? This is a question worthy of further investigation.

ENDNOTES

The author thanks the Andrew Mellon Foundation for supporting his research on Catholic schooling.

1. Sander and Krautmann (1995) present results that are similar in several respects to those of Evans and Schwab.
2. See Coleman and Hoffer (1987, p. 131). However, given Coleman and Hoffer's method of presenting results, it is difficult to determine whether or not the differences in gaps across groups are statistically significant.
3. In their analyses, which correct for selection bias, Evans and Schwab (1995) also report slightly larger Catholic school effects for blacks than for whites. However, the differences are small and statistically insignificant.
4. A great portion of the debate took place in three issues of the *Sociology of Education* in 1982, 1983, and 1985. My references contain several articles from these issues. Murnane (1981) provides a review of much of the literature that deals with the original Coleman, Kilgore, and Hoffer report. Heckman and Neal (1996) also review this literature.
5. The exact magnitude of the achievement gains varies with grade level and econometric specification. However, the estimated effects are always large for urban minorities. As an example, a standard analysis of the tenth-grade math scores yields a 7 percent gain in achievement from Catholic schooling for blacks in cities.

REFERENCES

- Alexander, Karl L., and Aaron M. Pallas. 1983. "Private Schools and Public Policy: New Evidence on Cognitive Achievement in Public and Private Schools." *SOCIOLOGY OF EDUCATION* 56: 170-81.
- . 1985. "School Sector and Cognitive Performance: When Is a Little a Little?" *SOCIOLOGY OF EDUCATION* 58: 115-28.
- Coleman, James, and Thomas Hoffer. 1987. *PUBLIC AND PRIVATE SCHOOLS: THE IMPACTS OF COMMUNITIES*. New York: Basic Books.
- Coleman, James, Sally Kilgore, and Thomas Hoffer. 1981. *PUBLIC AND PRIVATE HIGH SCHOOLS*. Washington, D.C.: National Center for Educational Statistics.
- Evans, William, and Robert Schwab. 1995. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *QUARTERLY JOURNAL OF ECONOMICS* 110: 947-74.
- Figlio, David N., and Joe A. Stone. 1997. "School Choice and Student Performance: Are Private Schools Really Better?" Institute for Research on Poverty Discussion Paper no. 1141-97.
- Goldberger, Arthur S., and Glenn G. Cain. 1982. "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer, and Kilgore Report." *SOCIOLOGY OF EDUCATION* 55: 103-22.
- Greene, Jay P., Paul E. Peterson, and Jiangtao Du. 1987. "Effectiveness of School Choice: The Milwaukee Experiment." Harvard University Program in Education Policy and Governance Paper no. 97-1.
- Heckman, James. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions in One Widely Used Estimator." *JOURNAL OF HUMAN RESOURCES* 32: 441-62.
- Heckman, James, and Derek Neal. 1996. "Coleman's Contributions to Education: Theory, Research Styles, and Empirical Research." In Jon Clark, ed., *JAMES S. COLEMAN*. London: Falmer Press.
- Heckman, James, and Richard Robb. 1985. "Evaluating the Impact of Treatments on Outcomes." In J. Heckman and B. Singer, eds., *LONGITUDINAL ANALYSIS OF LABOR MARKET DATA*. Cambridge: Cambridge University Press.
- Hill, Paul T., Gail E. Foster, and Tamar Gendler. 1990. "High Schools with Character." RAND R-3944-RC.
- Murnane, Richard J. 1981. "Evidence, Analysis, and Unanswered Questions." *HARVARD EDUCATIONAL REVIEW* 51: 483-9.
- Murnane, Richard J., Stuart Newstead, and Randall J. Olsen. 1985. "Comparing Public and Private Schools: The Puzzling Role of Selection Bias." *JOURNAL OF BUSINESS AND ECONOMIC STATISTICS* 3: 23-35.
- Neal, Derek. 1997. "The Effect of Catholic Secondary Schooling on Educational Attainment." *JOURNAL OF LABOR ECONOMICS* 15: 98-123.
- Nechyba, Thomas J. 1997. "Public School Finance in a General Equilibrium Tiebout World." Unpublished paper, Stanford University.
- Noell, John. 1982. "Public and Catholic Schools: A Reanalysis of Public and Private Schools." *SOCIOLOGY OF EDUCATION* 55: 123-32.
- Rouse, Cecilia Elena. 1998. "Schools and Student Achievement: More Evidence from the Milwaukee Parental Choice Program." Federal Reserve Bank of New York *ECONOMIC POLICY REVIEW* 4, no. 1.
- . Forthcoming. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *QUARTERLY JOURNAL OF ECONOMICS*.
- Sander, William. 1997. "Catholic High Schools and Rural Academic Achievement." *AMERICAN JOURNAL OF AGRICULTURAL ECONOMICS* 79: 1-12.
- Sander, William, and Anthony Krautmann. 1995. "Catholic Schools, Dropout Rates, and Educational Attainment." *ECONOMIC INQUIRY* 33: 217-33.
- Willms, J. Douglas. 1985. "Catholic-School Effects on Academic Achievement: New Evidence From the High School and Beyond Follow-Up Study." *SOCIOLOGY OF EDUCATION* 58: 98-114.
- Witte, John F. 1992. "Private School versus Public School Achievement: Are There Findings that Should Affect the Educational Choice Debate?" *ECONOMICS OF EDUCATION REVIEW* 11: 371-94.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

Measuring the Value of Better Schools

Sandra E. Black

In 1993, spending on education represented 28 percent of state and local government expenditures and almost 14 percent of total government expenditures in the United States.¹ The tremendous resources devoted to education in this country underscore the need to identify the tools and programs that yield the greatest return on our investment. Policymakers have sought to improve schools in a variety of ways, ranging from increasing per pupil expenditures or teacher salaries to creating programs that send inner-city students to suburban schools. How, then, do we assess the cost-effectiveness of specific initiatives and programs?

The first step is to measure the value of better schools. The goal is to develop a sound method of quantifying how investments in educational quality relate to outcomes. Once we are able to put a dollar value on improvements in school quality, we can compare policy alternatives.

Sandra E. Black is an economist at the Federal Reserve Bank of New York.

In this paper, I examine two methods of measuring the value of better schools. One involves following individuals over time to determine how the quality of their schooling affects outcomes later in their lives; the other involves calculating parental valuation of better schools today. I review the benefits and limitations of the two methods, then briefly consider how these methods might be used in evaluating policies. At the end of the paper, I note some uncertainties affecting research in the field and outline directions for future research.

THE RELATIONSHIP BETWEEN INPUTS AND OUTPUTS

Determining the value of better schools is more difficult than it seems. The effects of better schools are fully realized only with the passage of time: the benefits of attending a very good elementary school, for example, may stretch over a lifetime. Nevertheless, assessing school value is easier if we break the task down into two stages—first, determining the relationship between inputs to a school and outputs, and second, determining the dollar value of this increased output (see figure on page 88).

Inputs to a school can be characterized by a number of variables. Teacher salaries, per pupil expenditures, and expenditures to achieve a particular student-teacher ratio are forms of financial input, which is what is most commonly understood by the notion of inputs. Also important, however, are the nonfinancial forms of input—the mix of students in the school (an indicator of overall peer quality), parental time and resources, and the quality of the administration. All of these variables—financial and nonfinancial—can be used by researchers as indicators of school quality.

School outputs can also be measured in a number of ways. Improvements in test scores are an indicator of school success. So are financial gains, such as higher wages later in life. Intermediate to these is educational attainment: students who attend better schools may stay in school longer.

Once we establish a relationship between inputs and outputs, we need to put a dollar value on the increased output.² In some cases, the work is already done for us: wages earned later in life are one measure of output that needs no quantification.³ But attaching a value to higher test scores or to longer stays in school is a more complicated undertaking. One way to do this is by calculating how much people are willing to pay for a home in a location that would allow their children to attend a better school—an approach that essentially measures the capitalization of better schools in house prices.

The next two sections examine in more detail how researchers use the concepts of input and output to develop techniques for measuring the value of better schools. The

first technique focuses on the relationship between inputs to schools and children’s outcomes later in life; the second looks at the links between inputs, outputs, and parental willingness to pay.

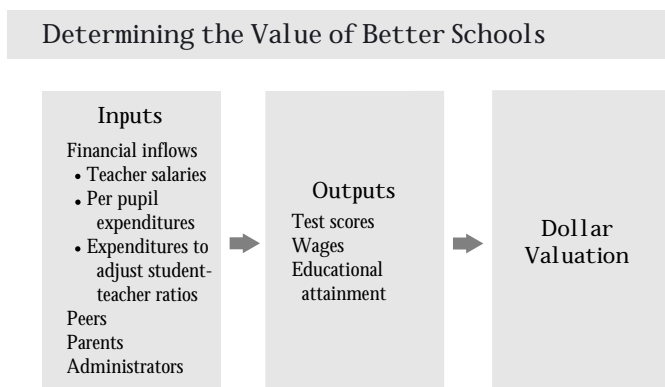
TRACKING INDIVIDUAL OUTCOMES

The advantage of following individuals over time is that you can look at a number of outcomes for each individual and assess the value of better schools in several different ways.⁴ For example, you can see how better schools affect wages, educational attainment, and job choice. Under this approach, the optimal experiment would be to compare two individuals who are identical in all respects (family background, innate ability, and so forth) except for the schools they attended. Any differences in outcomes could then be attributed to differences in school quality.

In practice, of course, we cannot compare two identical individuals. In addition, educational quality is only one of many determinants of an individual’s wage or educational attainment, and we have imperfect controls for the other determinants, such as family background. Therefore, any relationship we observe between outcomes and school quality may be tainted by “omitted variable bias” if we overlook, or cannot control for, differences in students’ backgrounds or innate abilities. The danger is that we will overstate the effect of school quality on individual outcomes because we cannot adjust adequately for the effects of these other factors.

Because of the difficulty of measuring the relationship between school inputs and individual outcomes, the conclusions reached in the literature vary significantly with the outcome measure used. Problems such as short panels of wage data, state-level rather than local information about school quality, and incomplete characterizations of family backgrounds plague these studies.

To date, much of the literature has focused on the relationship between the financial inputs to schools and the wages earned by students later in life.⁵ When following individuals over time, it makes sense to put a dollar value on better schools by calculating the influence of better schools on wages.⁶ Using census data, Card and Krueger (1992a) estimate the relationship between the wage



return to a year of school and the quality of that year of school, where the measures of quality—the student-teacher ratio, the term length in days, and the relative teacher wage in the students’ state of birth—reflect the financial resources available to schools in that state.⁷ The authors find a significant relationship, suggesting that financial inputs do matter. In a summary article (Card and Krueger 1996), they suggest that a 10 percent increase in school spending is associated with a 1 to 2 percent increase in earnings for students.

This result is contradicted in work by Betts (1995). Using the National Longitudinal Survey of Youth (NLSY) data set, Betts finds no significant relationship between the financial resources available to *individual* schools and students’ earnings. However, when he substitutes state-level measures of school quality for measures at the individual-school level, he, like Card and Krueger, finds a significant relationship, suggesting that Card and Krueger are capturing state, rather than individual-school, effects. A criticism of Betts’ work, however, is that the data contain information about wages early in life, and one might argue that the effects of school quality are not realized until later.

VALUATION AS CAPITALIZED IN HOUSING PRICES

Given the controversy surrounding the results of this approach, one is left looking for other methods of measuring the value of better schools. A second approach involves determining how much people are willing to pay for better schools. We can infer this value by examining how much more people pay for houses located in areas with better schools.⁸ Although this methodology may seem indirect—it measures the value of better schools to parents, not the value to the child receiving the education—it has the advantage of putting a dollar value on current school quality, as opposed to school quality from many years earlier. Another advantage of the approach, as we will see below, is that it allows the analyst to minimize the potential for omitted variable bias.

Calculating the value of better schools this way calls for a two-step procedure: the analyst first evaluates the

relationship between inputs—most often, financial inputs to the school—and a measure of output—typically, the average test scores for that school. The analyst then derives the willingness to pay for higher test scores by examining how school test scores are capitalized in housing prices.⁹ Parental willingness to pay serves as a measure in dollar terms of the benefit of higher student test scores;¹⁰ this benefit can then be compared with the costs of different educational programs to determine each program’s cost-effectiveness.

The literature examining the relationship between financial inputs to schools and test scores has generally followed the methodology used to assess the relationship between financial inputs and wages.¹¹ Hanushek (1986, 1996a, 1996b) finds little evidence to suggest that financial inputs to a school have any significant effect on student test scores. Even when evaluating the large number of existing studies on this topic—some of which claim to find a link between school financial resources and test scores—he still concludes that the relationship does not exist.

Using the same studies as Hanushek, however, Hedges, Laine, and Greenwald (1994) come to quite a different conclusion. Applying a meta-analysis technique to evaluate the existing body of research, they find a significant relationship between financial inputs and achievement, such that a \$500 increase in average per pupil expenditures (approximately a 10 percent increase) leads to a .7 standard deviation increase in student achievement.

Other researchers agree with Hanushek that the evidence of a relationship in existing studies is scant, but claim that data limitations are responsible. Ferguson (1991), for example, uses more detailed Texas data to show that better teachers lead to improved student performance. His conclusion is not inconsistent with Hanushek’s position, however; Hanushek acknowledges that a relationship exists between the quality of schools and student achievement, but argues that this relationship cannot be explained by the measurable financial inputs to the schools.

Although the relationship between financial inputs and outputs is unclear, it is still important to have an understanding of the value of better schools. Educators and

policymakers have nonfinancial means of upgrading school quality. Given that improvements are possible, we can get a relatively clean estimate of parental willingness to pay for better schools by looking at the increase in house prices, or capitalization, that is associated with schools whose students, on average, score higher on tests.

When measuring the value of schools through the capitalization of test scores in house prices, we would like to compare two houses that are identical except that the children in one house attend a better school than the children in the other house. In that case, any difference in house prices can be attributed to differences in school quality. In practice, however, measurement is complicated by the difficulty of isolating school quality effects from better neighborhood effects. Since better schools tend to be located in better neighborhoods, ordinary hedonic housing price regressions of the form

$$\log(\text{price}_{ijk}) = \alpha + X_{ijk}\beta + Z_j\delta + S_k\theta + \varepsilon_{ijk},$$

where price_{ijk} represents the selling price of house i in neighborhood j in school district k , X_{ijk} is a vector of house-level characteristics, Z_j is a vector of neighborhood characteristics, and S_k represents school quality characteristics, may lead to an overstated valuation of better schools if the available data do not provide a complete characterization of the neighborhood studied.

Early willingness-to-pay studies tended to look at large, heterogeneous areas in measuring the value of better schools.¹² Because these studies did not control adequately for neighborhood differences, they were very susceptible to omitted variable bias. To avoid such bias, more recent work has attempted to control for neighborhood differences by focusing on increasingly localized areas.

To understand how the literature has evolved, consider first a study by Jud and Watts (1981) that examined one school district—Charlotte, North Carolina. The authors found a significant and meaningful relationship between house prices and the average scores of the schools' third-grade students on the state test of reading skills. However, since the geographic area under study was quite large, the houses compared may have been in entirely different neighborhoods. Because the authors

controlled for only a limited number of neighborhood characteristics, their comparison may not have been valid.

Other work has attempted to correct for this problem by comparing houses in smaller geographic areas. Work by Hayes and Taylor (1996) and Clotfelter (1975), although not specifically focusing on the valuation of schools, looked at houses within the same school attendance district, where the attendance district is the geographical area that defines which school *within* a school district a child will attend.¹³ Because the authors examined a smaller geographic area, the variation in neighborhoods across houses being compared should have been less.

Even with the focus narrowed to attendance districts, however, omitted neighborhood differences might still bias estimates of the value of higher student achievement. Two houses at opposite ends of an attendance district may be situated in very different neighborhoods. In a recent study (Black 1997), I address this problem by examining an even more localized area. Specifically, I compare the price of houses on opposite sides of elementary school attendance district boundaries in suburban Boston. Such a strategy, in its purest form, would restrict the area of the houses being compared to the point where there was no variation in neighborhoods. Imagine, for example, two houses on opposite sides of a street that forms the attendance district boundary. Children in the house on one side of the street attend a different school from the children in the house on the other side of the street, but the neighborhood is unquestionably the same. In such a case, any difference in prices would be attributable to differences in school quality.¹⁴ Because attendance district boundaries are within school districts and within a city, variations in property tax rates would be eliminated.

My study gets very close to this ideal comparison. Although data limitations prevent me from looking at houses on opposite sides of the same street, I am able to limit my sample to houses located within a relatively short distance of the boundary. I then narrow the sample to houses located closer and closer to the attendance district boundaries in order to minimize the likelihood that omitted neighborhood characteristics are driving the

results of the estimation. The final comparison I draw is of houses within .15 miles of an attendance district boundary.

The estimated equation is:

$$\log(\text{price}_{iab}) = \alpha + X_{iab}\beta + K_b\Phi + \gamma \text{test}_a + \varepsilon_{iab},$$

where price_{iab} is the selling price of house i in attendance district a near boundary b , X_{iab} is a vector of house characteristics, K_b is a vector of boundary dummies, and test_a is the measure of school quality assessed at the individual school level.

Using this boundary fixed-effects technique, I find that substantial omitted variable bias exists when one relies on the standard controls for neighborhood characteristics. Significantly, my estimate of the value of better schools is only about half of the normal hedonic housing price estimate. Controlling for neighborhoods and school financial inputs, I find that a 5 percent increase in elementary school test scores (a change of approximately one standard deviation in the observed data) leads to an increase in house prices of approximately 2.1 percent, or \$4,000 at the mean house price of the sample. From another perspective, a movement from the twenty-fifth-percentile school in the sample as ranked by test scores to the seventy-fifth-percentile school results in a 2.9 percent increase in house prices, or \$5,500 at the mean house price.¹⁵

How can we be sure that this procedure actually estimates the value of better schools? The study includes a number of specification checks, including checks to determine whether the attendance district boundaries chosen represent neighborhood divisions. For example, I eliminate any boundaries that could be major roads, and I control for measurable neighborhood characteristics (evaluated at the level of the census block group). A particularly compelling check involves comparing the results for one- and two-bedroom houses with the results for houses containing three or more bedrooms. One would expect individuals who live in houses with three or more bedrooms to be willing to pay more for better schools than people in smaller houses because they are more likely to have children. The study very clearly confirms this expectation.

This type of estimation provides a measure of the value of higher test scores that is remarkably free from

omitted variable bias. The analyst who follows students over time to determine how the quality of their schooling relates to their wages later in life cannot easily control for influences and events that affect students outside of school. In contrast, the analyst who looks within neighborhoods at the relationship between school quality and house prices can significantly reduce the number of omitted variables.

This estimation technique does, however, have some limitations. Studies employing the technique must focus on small localities—in the case of my study, suburban Boston—and, consequently, generalizing results to a wider area requires strong assumptions. In addition, in order to look at attendance district boundaries, such studies must look within school districts, which is the level at which school inputs such as spending are determined. As a result, the variation in school spending is significantly reduced, and calculations are based on differences in test scores that are attributable for the most part to the nonfinancial inputs to a school. Therefore, the variable of interest reflects differences in teacher quality, administrator quality, parental involvement, and school composition (peer effects). That is not to say that the value of high test scores will change when financial inputs vary; we do not currently have enough information to determine how differences in expenditures would affect the results.

CONCLUSION

At present, there is no perfect way to measure the value of better schools. For those charged with evaluating school policy, the best approach would be to combine the information acquired using both techniques explored in this paper and to draw inferences using all available evidence.

Current school quality evidence indicates that increased spending will not automatically improve student outcomes; this finding suggests that we should look at other ways to improve schools. Evidence from Black (1997) confirms the value of raising test scores, particularly through parental and administrative involvement and the influence of school peers.

How can we use these research findings when evaluating government policies? One way to apply my estimate of the value of higher test scores is in the analysis of policies involving student integration. Consider, for example, reforms such as the METCO program in Boston, which sends a few students from poorer urban neighborhoods to wealthier suburban schools. Suppose that a student is transferred from an inner-city Boston elementary school where test scores average 21.27 to an elementary school in the wealthier suburb of Belmont where test scores average 30.67. We can measure the benefit of this test score improvement by calculating the percentage difference in house prices for two equivalent houses in the same neighborhood but in different attendance districts. Using the estimate obtained in Black (1997), we find that a house would appreciate 14 percent in value if the family residing there acquired the right to send their child to the elementary school in Belmont instead of the one in Boston. When evaluated at the mean housing price in my sample, this number in dollar terms is approximately \$25,660.¹⁶ We can then compare this

benefit with the costs of implementing the program and thereby evaluate the program's cost-effectiveness.

In the end, it is important to think about what we can and cannot say. We can say that parents are willing to pay more for better test scores, although we do not have a clear understanding of the relationship between spending on inputs and test scores. We can say that evidence suggests a relationship between school inputs and the wages earned by students later in life, but we would be hard pressed to assign an actual dollar value to the school inputs. Our uncertainty suggests a need for communication with educators and other policymakers. Their knowledge and experience can help researchers to identify the best ways to improve student performance—through hiring better teachers, boosting parental and administrative involvement, creating the optimal mix of students in a classroom, or increasing the efficiency with which schools use financial resources. Finally, we must continue to improve our methods of evaluation by collecting relevant data and seeking an experimental design that eliminates biases in our estimates.

ENDNOTES

The author thanks Carol Rapaport, Joe Tracy, and Joyce Black for helpful discussions and comments. Colleen Sellers provided outstanding research assistance.

1. See U.S. Department of Commerce, Bureau of the Census (1996).
2. In this paper, I focus on the private, rather than the social, valuation of better schools. Those who believe that education is a public good would argue that the private valuation understates the true value of education. Two studies exploring these issues are Kane and Rouse (1995) and Rauch (1993).
3. Note that while student wage increases are one obvious way of putting a dollar value on better schools, this valuation does not incorporate other, nonmonetary benefits such as a more pleasant work environment, a more interesting job, and the like.
4. See Burtless (1996) for a review of the studies that take this approach.
5. Because the literature is vast, this paper can only highlight a few of the more representative studies.
6. A primary source of evidence supporting the positive relationship between school quality and earnings is work that relates relative changes in school quality for one group to changes in relative wages for that group. For example, a large literature focuses on changes in school quality for blacks and the subsequent shift in black-white earnings differentials. Card and Krueger (1992b) find that between 1960 and 1980, improvements in the relative quality of black schools explain 20 percent of the narrowing of the black-white earnings gap.
7. Essentially, Card and Krueger estimate individual-level wage equations and allow for state-specific intercepts and education slopes. They then take these state-specific education coefficients and regress them on state school quality averages and other state-level data.
8. Another way to calculate people's valuation of a good is to ask them directly how much they value the particular good. This contingent valuation approach is widely used in the environmental literature (see Cummings, Brookshire, and Schulze [1986]). However, because the quality of the answers obtained through this method is often doubtful, most economists prefer to use a market-determined estimate of valuation.
9. Early work by Rosen and Fullerton (1977) shows that test scores perform better than per pupil expenditures as a measure of school quality in property value equations. More recently, work by Hayes and Taylor (1996) suggests that parents actually do focus on the value added of a school and not, as one might expect, the inputs to the school.
10. Although parental willingness to pay is equated with the value of higher test scores, it may in fact also be picking up the value of other things that are correlated with higher test scores.
11. Again, see Burtless (1996) for a review.
12. See work by Kain and Quigley (1975) for an example of this literature.
13. Hayes and Taylor focus more specifically on whether parents use test scores or inputs to the school as measures of school quality. The authors find that property values reflect student test scores but not school expenditures, and they conclude that the relationship between test scores and property values arises from an underlying relationship between property values and the marginal effects of schools. Clotfelter uses attendance districts to look at the effect of school desegregation on housing prices.
14. In the construction of the data set, boundaries that represented clear neighborhood divisions such as railroad tracks or parks were excluded from the sample.
15. These estimates are also robust to a number of specification tests.
16. The calculation would be $(30.67-21.27)*.015 = 9.4*.015 = .14$, where .015 is the coefficient on the elementary school test score in the hedonic housing price regression estimated in Black (1997). Note that this policy application requires strong out-of-sample assumptions.

REFERENCES

- Betts, Julian R.* 1995. "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth." *REVIEW OF ECONOMICS AND STATISTICS* 77 (May): 231-50.
- Black, Sandra E.* 1997. "Do Better Schools Matter? Parental Valuation of Elementary Education." Federal Reserve Bank of New York Research Paper no. 9729.
- Burtless, Gary.* 1996. "Introduction and Summary." In Gary Burtless, ed., *DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS*. Washington, D.C.: Brookings Institution.
- Card, David, and Alan B. Krueger.* 1992a. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *JOURNAL OF POLITICAL ECONOMY* 100 (February): 1-40.
- . 1992b. "School Quality and Black-White Relative Earnings: A Direct Assessment." *QUARTERLY JOURNAL OF ECONOMICS* 107 (February): 151-200.
- . 1996. "Labor Market Effects of School Quality: Theory and Evidence." NBER Working Paper no. 5450.
- Clotfelter, Charles T.* 1975. "The Effect of School Desegregation on Housing Prices." *REVIEW OF ECONOMIC STATISTICS* 57, no. 4 (November): 446-51.
- Cummings, Ronald G., David S. Brookshire, and William D. Schulze, eds.* 1986. *VALUING ENVIRONMENTAL GOODS: AN ASSESSMENT OF THE CONTINGENT VALUE METHOD*. Totowa, N.J.: Rowman and Littlefield.
- Ferguson, Ronald.* 1991. "Paying for Public Education: New Evidence on How and Why Money Matters." *HARVARD JOURNAL ON LEGISLATION* 28 (summer): 465-98.
- Hanushek, Eric A.* 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *JOURNAL OF ECONOMIC LITERATURE* 24 (September): 1141-77.
- . 1996a. "Measuring Investment in Education." *JOURNAL OF ECONOMIC PERSPECTIVES* 10 (fall): 9-30.
- . 1996b. "School Resources and Student Performance." In Gary Burtless, ed., *DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS*, 43-73. Washington, D.C.: Brookings Institution.
- Hayes, Kathy, and Lori Taylor.* 1996. "Neighborhood School Characteristics: What Signals Quality to Homebuyers?" Federal Reserve Bank of Dallas *ECONOMIC REVIEW*, fourth quarter: 2-9.
- Hedges, Larry V., Richard Laine, and Rob Greenwald.* 1994. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *EDUCATIONAL RESEARCHER* 23, no. 3 (April): 5-14.
- Jud, G. Donald, and James Watts.* 1981. "Schools and Housing Values." *LAND ECONOMICS* 56, no. 2: 459-70.
- Kain, John, and John M. Quigley.* 1975. *HOUSING MARKETS AND RACIAL DISCRIMINATION: A MICROECONOMIC ANALYSIS*. New York: NBER.
- Kane, Thomas J., and Cecilia Elena Rouse.* 1995. "Labor-Market Returns to Two- and Four-Year College." *AMERICAN ECONOMIC REVIEW* 85, no. 3: 600-14.
- Rauch, James E.* 1993. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *JOURNAL OF URBAN ECONOMICS* 34.
- Rosen, Harvey S., and David J. Fullerton.* 1977. "A Note on Local Tax Rates, Public Benefit Levels, and Property Values." *JOURNAL OF POLITICAL ECONOMY* 85, no. 2.
- U.S. Department of Commerce. Bureau of the Census.* 1996. *STATISTICAL ABSTRACT OF THE UNITED STATES*. Table 472, p. 298.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

Summary of Floor Discussion

James Orr

The discussion following the third-session papers initially centered on Sandra Black's finding that parents are willing to pay more for a house in a school district where students' average test scores are relatively high. Richard Murnane noted that the source of the differences in test scores among areas is an important issue for school policy. He then asked whether Black's findings are consistent with the notion that peer groups matter for school quality, suggesting that higher test scores result from interaction among children who care about their education and that some parents are willing to pay extra to expose their children to similarly motivated students. Black answered that her findings are consistent with this notion, but stressed that the paper did not address the source of the higher test scores. Derek Neal pointed out that differences in housing prices are likely to persist only in relatively crowded urban or suburban areas, where there is limited room for an expansion of higher priced housing. Black responded that she was looking at relatively densely populated suburbs, so the supply of housing could be considered constant.

The discussion then turned to the policy implications of Derek Neal's paper, which outlined the gains made by urban minority students who attend private, mainly Catholic, secondary schools. Amy Schwartz observed that the problems with schools are not wholesale ones that are found equally in urban and suburban areas; Neal had focused correctly on the problems of education in urban areas. For example, a policy that lowers average class size,

Schwartz said, could well be more effective if it was targeted at New York City's relatively large class sizes rather than a suburban district's already modest class sizes. She then reminded the group that by providing financial incentives to parents, we empower them—rather than teachers or others—to make decisions about what constitutes a good school. Schwartz also posed the question whether suburban commuters would choose to live in cities if they had access to good public schools.

Next, Joseph Viteritti remarked that the biggest gap in public education occurs in the inner city, which is also where Catholic schools have demonstrated the biggest effect. Several ideas were put forth by participants to explain why Catholic secondary schools are outperforming their public school counterparts in the inner city: Catholic schools set higher performance standards and have high expectations of each child; they emphasize basics in the curriculum; they favor a nonbureaucratic structure. Viteritti noted that the success rate in Catholic elementary schools in the inner city is likely even higher than in the secondary schools. He stated that an effective school choice program should be needs-based and should give public schools greater autonomy, similar to the autonomy found in many charter schools. Neal pointed out that the disparity between white and minority high school graduation rates is greatest in the larger urban areas. He expressed some skepticism about how much of any additional resources given to school systems in these areas would actually reach the

classroom, arguing that the size of the educational bureaucracy affects allocation to students.

Maureen O'Brien then drew a distinction between the roles of school principals in the public and private school systems in New York City. She stressed that the principals are the chief agents of change in the Catholic schools and are heavily involved in activities such as curriculum planning and teacher promotions, while they do not have a similar role in public schools. Caroline Hoxby cited some statistical support for O'Brien's assertion of the importance of good principals and noted that an issue that

policymakers should address is how to guarantee a supply of good principals and then empower them. Neal agreed that the structure of public school systems in many large urban areas often does not give principals the necessary authority to reward teachers doing a good job or to take action against those doing a poor job.

The discussion closed on a cautionary note: changes in reading scores over time may well be a better measure of a school's performance than the overall level of reading scores.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

The Two-Legged Stool: The Neglected Role of Educational Standards in Improving America's Public Schools

Julian R. Betts

Two of the most important reforms to American public schooling in this century have been an increase in the minimum school-leaving age and a dramatic increase in expenditures per pupil. The former reform has generally been hailed as a success, given evidence that an extra year of schooling significantly boosts students' earnings later in life. However, evidence on the effectiveness of the trend toward higher spending per pupil, smaller class sizes, and more highly educated and trained teachers is much more mixed. A host of studies on the link between school finances and test scores has not shown a systematic link between spending and achievement. Another set of studies tests whether higher school spending leads to higher earnings for students later in life. The findings in this body of work are also mixed: even the most optimistic

results suggest a very low rate of return to increased school expenditures.

Given the central role of public schools in preparing younger generations for the workforce and the sizable expenditures devoted to public schools, it becomes important to ask whether other reforms might increase the effectiveness of public schooling. The premise of this paper is that educational standards are a key element in school reform. The paper suggests that the twin policies of higher school-leaving ages and higher spending would have been much more effective if accompanied by systematic increases in educational standards. In a sense, these two policies form a two-legged stool. Increasing the number of years that students must spend in school and increasing spending per pupil—without at the same time stipulating what subject matter students should have mastered by each age—are unlikely to maximize the rate at which students learn. The missing “leg” in these past reforms is a set of academic standards against which both students *and*

Julian R. Betts is a professor of economics at the University of California, San Diego.

schools are measured. Instead of allowing students to drop out of school merely because they have reached the age of sixteen, why not require students who wish to leave school before graduation to pass a set of minimum competency tests? Similarly, does it make sense to increase funding for public schools without at the same time requiring the schools to document that they have translated these additional expenditures into improved student achievement? In both cases, the need for a clear set of curriculum standards, backed by testing, is clear.

The central point is simple: far and away the most important determinant of how quickly students learn is the effort of students themselves. It follows that an increase in schools' expectations of students could have important effects on the quality of public schooling. By establishing a rigorous set of educational standards, schools can create a set of incentives and rewards to promote student learning. Higher standards can increase the effectiveness of school-leaving policies. Instead of simply allowing students to drop out at age sixteen, schools might require an exit exam. In this way, a minimal level of academic achievement, rather than age alone, would become the prerequisite for dropping out of school. Similarly, reductions in class size might become more effective if, at the same time, schools increased the standards that students at each grade level are expected to meet.

The next section briefly outlines the history of past reforms related to the minimum school-leaving age and spending per pupil. The subsequent section outlines ways in which schools can heighten their expectations of students. In practice, "higher expectations" can come in the form of additional homework, the development of curriculum standards in conjunction with an assessment of students' progress in mastering the curriculum, stricter grade-promotion policies, and more stringent grading of students. Later sections then analyze whether such policies to promote higher educational standards work in practice. The penultimate section delves into practical and political issues that can afflict a school administration when it tries to increase standards, and suggests solutions to some of these problems.

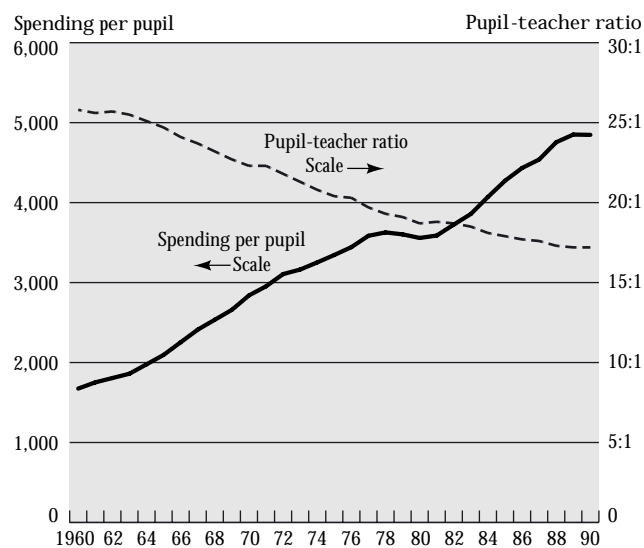
A REVIEW OF THE IMPACT OF PAST INCREASES IN THE SCHOOL-LEAVING AGE AND IN SPENDING PER PUPIL

Between 1960 and 1990, real spending per pupil and the teacher-pupil ratio increased dramatically (Chart 1). (Expenditures per pupil are expressed in 1990-91 prices, and so account for inflation.) Betts (1995a), among others, documents that there has been a strong trend toward increased spending per pupil throughout the century in the United States. This trend represents one of the most important changes in the recent history of public schooling.

A second important reform in American public schooling has been increases in the school-leaving age. Lang and Kropp (1986) document that over time the number of states without a school-leaving age has dropped, and the average school-leaving age has risen considerably (Chart 2). Most of the changes in the school-leaving age occurred in the first half of the century. (However, Lang and Kropp note that by 1965 two states—Mississippi and South Carolina—had abolished their compulsory attendance laws.)

Chart 1

Trends in Real Spending per Pupil and in the Pupil-Teacher Ratio



Source: National Center for Education Statistics (1991).

Note: Spending per pupil is in 1990-91 prices.

What benefits, then, have resulted from these two important transformations in American public schooling? Consider first the impact of additional school spending. For over three decades, social scientists have examined the link between school expenditures and student success. Most of this effort has modeled test scores as a function of spending per pupil, the teacher-pupil ratio, and other measures of school inputs. Since other contributors to this conference volume address this literature, I will not review it in detail here. It suffices to mention that in a recent review, Hanushek (1996) found that of 163 estimates of how spending per pupil affects student performance, only 27 percent found a positive and significant relationship. Similarly, of 277 reported estimates of the impact of the teacher-pupil ratio on student performance, only 15 percent found a positive and significant link, while 13 percent reported a negative and significant link. These figures do not mean that money never matters. Instead, they suggest that the relationship, if positive, is a rather tenuous one. In American schools, at least as they have operated in the past, spending has not had large or systematic effects on student achievement.

The conclusion drawn from the statistical research is supported by aggregate trends in school spending and in student achievement. Chart 1 shows trends in current expenditures per pupil and the pupil-teacher ratio in American public schools between 1960 and 1990. By both measures, the financial resources spent on public school students have risen markedly over the last three decades. Yet during the same period, student achievement has hardly changed, and by one measure it may even have fallen. Test scores on the National Assessment of Educational Progress, a test given to a random sample of students in various grades since the early 1970s, have changed little over the 1970s and 1980s. Trends in the Scholastic Aptitude Test show a sharp decline in the late 1960s, a more gradual decline during the 1970s, and a partial recovery since then (see, for instance, Hanushek [1996]).

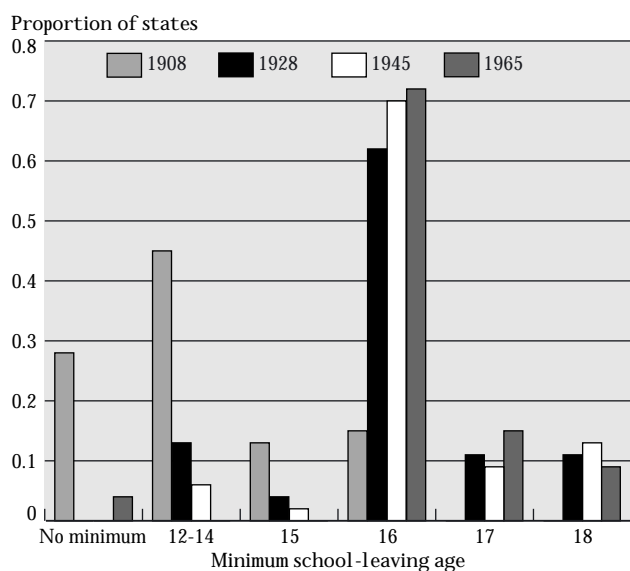
From an economist's perspective, the underlying goal of education is to prepare younger generations for success in adult life. A major determinant of adult success is earnings. Betts (1996a) surveys the literature that asks whether school spending affects students' earnings later in life, even if there appears to be little relationship between school resources and test scores. Quite a number of studies have found a relationship between adult males' earnings and school resources in their state of birth. But the literature is by no means unanimous. Work by Betts (1995a) and Grogger (1996), among others, shows that when school resources are measured at the school actually attended, the results are much more consistent with the test score literature: the impact of school inputs on earnings is not statistically significant.

More to the point, the estimated effect of raising school spending on students' subsequent earnings is extremely small. This is true whether one measures school resources at the school actually attended or the district attended, or whether one instead uses the person's state of birth to create a rough proxy for school resources.

The small impact of school spending on earnings stands in stark contrast to a voluminous literature documenting that a person's years of schooling are strongly related to subsequent earnings. Mincer's (1974) seminal contribution estimates that one additional year

Chart 2

Distribution of Minimum School-Leaving Ages across States, by Year



Source: Lang and Kropp (1986).

of schooling typically increases wages by about 7 to 10 percent. Psacharopoulos (1985), in a review of the literature, estimates that in developed countries a year of schooling typically is associated with a 9 percent rise in earnings. He also reports much higher returns to years of schooling in developing countries—a finding that invites us to ask what might be done in developed countries, such as the United States, to increase the payoff to a year of schooling.

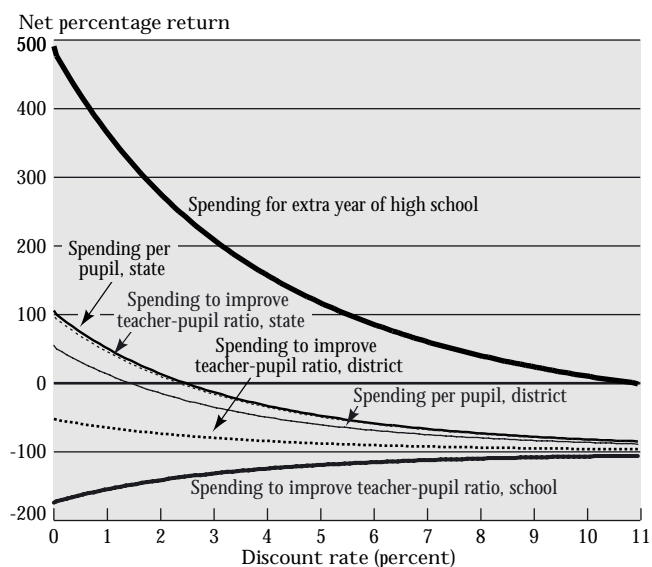
Simple calculations suggest that the economic returns to raising the school-leaving age are fairly substantial, but that the economic returns to increasing school spending are relatively meager. Consider first the rate of return to increasing spending per pupil or increasing the teacher-pupil ratio. It is important to discount future costs and benefits when calculating the net benefits from spending on a given school input, because all the costs are incurred in the early years of a person's life, while the benefits of higher wages accrue much later. The following calculations discount all costs and benefits to the year in which a student is in kindergarten. I assume that a 1 percent increase in spending per student is made in every school year between kindergarten and grade 12. To calculate the resulting increase in the discounted value of lifetime earnings, I multiply the predicted percentage wage gain, taken from the elasticities reported in Betts (1996a), by the discounted value of earnings for the average American male worker between the ages of nineteen and sixty-four. The calculation of the present discounted value of earnings uses the actual profile of earnings by age for male American workers, obtained using weighted earnings data from the March 1993 Current Population Survey tape. This tape contains information on annual earnings in 1992. The average annual earnings of male workers obtained from the Current Population Survey, taken as a simple mean across all ages from nineteen to sixty-four, was \$22,737.¹ Using this information, it is possible to calculate the net percentage return to an investment in school spending. This is calculated as the net return (wage gains minus the costs), divided by the costs, and expressed in percentage terms.

Similarly, one can calculate the net percentage return to an extra year spent in high school. Following Betts (1996a), assume conservatively that if we required a student who had dropped out of high school after grade 11 to remain in school for another year, the additional human capital he acquired would increase his earnings by 7.5 percent. This must be balanced against the cost of an extra year of schooling, which is wages forgone plus average spending per pupil in public schools, the latter of which is approximately \$5,000.² By calculating the estimated percentage wage gains and the initial wages lost from staying in school, again using the March 1993 survey, we can obtain different net returns on the initial investment for different "interest rates" or "discount rates."

The net percentage return to different types of educational expenditures is plotted in Chart 3. Note that as the discount rate rises, the net percentage return in general falls, because the given educational expenditure leads to higher predicted earnings for the student, but only later in life. As the discount rate rises, the present value of these wage gains declines.³

Chart 3

Net Percentage Return to Given Type of Spending versus Discount Rate



Source: Author's calculations, based on March 1993 Current Population Survey data.

The results are remarkable. The net percentage returns to requiring an extra year of high school are much higher than the returns to increasing school spending or to increasing the teacher-pupil ratio. As previously mentioned, when researchers have measured the actual school resources in a person's school district or school, the estimated effect of school spending is found to be smaller than when researchers instead use spending in the person's state of birth as a proxy. But even the most optimistic results—those that use the state-level data—suggest far smaller returns to increased spending per pupil or teacher-pupil ratios than are obtained when the student stays in school an extra year. Furthermore, in the district-level or school-level studies of the teacher-pupil ratio, even at a discount rate of 0 percent the initial costs of hiring more teachers are never recouped by subsequent increases in students' wages.

What is to explain the huge discrepancy between the returns to staying in school an extra year and the returns to increasing spending per pupil or increased teacher-pupil ratios? One possibility is that the returns to a year of education do not reflect true productivity gains. Instead, these returns might arise because of unobserved differences in ability between the highly educated and the less well educated. In other words, talented students might obtain more education merely to signal their ability to employers. If this "signaling" theory is true, then the observed returns to increasing the school-leaving age would be largely illusory.

Yet compiling evidence in favor of the idea that education merely signals a worker's productivity, rather than adding to productivity, has proved quite difficult. Lang and Kropp (1986) indirectly test for signaling. When a state raises its school-leaving age to, say, sixteen, then if students obtain education in order to signal their ability to employers, some seventeen- and eighteen-year-old students who would formerly have dropped out at age sixteen now begin to stay in school longer. They do this to "separate" themselves from their less able classmates. Lang and Kropp present some evidence that this has occurred historically. But numerous studies that have attempted to control for unobserved ability more directly have typically found that a year of schooling truly does increase productivity (see, for instance, Willis and Rosen [1979]). Numerous studies of

twins have attempted to sort out the true impact of a year of schooling on an individual's wage by comparing twins who obtained different levels of education. These studies have typically found that, if anything, the true productivity gain that results from an extra year of school is even higher than the 7.5 percent return that I used in the above analysis (see, for instance, Ashenfelter and Krueger [1994]). If these estimates are accurate, they suggest that increasing the school-leaving age would have substantial effects on the earnings of the affected students.

The best summary of the existing evidence is that, in the United States, the returns to increased spending per pupil or to increased teacher-pupil ratios have been extremely modest. The returns to an extra year of high school are much higher. However, as shown in Chart 3, even in this case the returns are not astronomically high: above a discount rate of 10.95 percent, the returns to additional education become negative. So, educational expenditures along the "extensive" margin (years of schooling) have been fairly productive, while expenditures along the "intensive" margin (spending per pupil) have had surprisingly small payoffs. This raises the question whether other aspects of public schools need to be changed in order to make financial inputs more effective than they have been in the past. The rest of this paper examines the proposition that the missing element in past reforms—the third leg of the stool—has been higher standards, higher expectations, and a higher degree of accountability in the nation's public schools.

A DEFINITION OF ACADEMIC STANDARDS AND ACADEMIC EXPECTATIONS, WITH EXAMPLES

It is easy to speak in general terms about higher standards and higher academic expectations for the nation's students. But what in practice does this mean? This section briefly outlines the necessary components for a variety of reforms that could represent a genuine shift toward higher standards.

CURRICULUM STANDARDS, TESTS OF ACHIEVEMENT, AND ACTIVE RESPONSES TO FAILURE

Perhaps the most often discussed way of increasing standards in schools is to strengthen the curriculum in

subjects such as mathematics, science, and English. As explained in the next section, many states have embarked on curriculum reform in one or more subjects. For curriculum reform to succeed, the subject content must be specified in detail, to ensure that all schools interpret the standard in the same way. A curriculum that states that “by grade 4, students will be able to express themselves well in written English” would not meet this criterion. It is so vague that it gives teachers little if any direction about what to teach. A far better approach would be to specify basic rules of grammar, spelling, and composition that should be mastered by that grade. The point of such a standard is not to tell teachers how to teach but to guide them as to what to teach, and when.

Creating a specific list of skills and knowledge that children should acquire in each grade is the first step. A curriculum standard by itself, however, is unlikely to improve schools substantially. It is also necessary to test students periodically to check whether they are meeting the standards set for each grade. Some states now have mandatory achievement tests in certain grades, but other states allow school districts to decide for themselves whether they want to test children. If so, the districts must choose among off-the-shelf tests or write a new test specific to their own curriculum. The need for a very specifically worded curriculum now becomes even more obvious: without a detailed curriculum, it will prove impossible to devise a test that gauges students’ academic progress.

The twin pillars of content standards, then, are a specifically worded curriculum and achievement tests that measure how well students are absorbing the prescribed curriculum.

It is certainly possible that testing based on content standards could improve the quality of schooling directly. Once parents become aware of any subject areas in which their children have fallen behind, they are likely to become more actively involved in their children’s schooling. Students themselves are likely to exert more effort during the academic year, knowing that at the end of the school year they will take a test that will inform their teachers and their parents about how much they have

learned. But there remains a distinct possibility that some students and some school administrations would disregard test results.

The investment in content standards and testing is likely to have a bigger payoff if there is something tangible at stake for both students and schools. The central question becomes, how, if at all, should a school react if a student does poorly on an achievement test? One obvious solution is for the school to devote additional resources—for example, tutoring or smaller classes—to such students. In other words, once a school system has developed a detailed curriculum and begins to test its students, it can direct additional spending to the students who need the most help. Systematic testing makes it possible to move away from a policy of improving schools through expensive systemwide increases in spending. This more focused approach to spending might make additional spending on schools much more effective than it has been in the past. To some extent, schools already direct more resources toward students whose achievement lags their grade level (see, for instance, Betts and Shkolnik [forthcoming]). However, the institution of a specific curriculum and regular testing based on this curriculum would allow more effective targeting of additional school expenditures toward children of relatively low achievement.

A second way in which schools can react when a student performs poorly on a test is to give the student a direct stake in his or her academic progress. A very long tradition in American schools has been to hold students back a year if they have not progressed sufficiently. Another approach is to require students who are lagging behind to attend summer school. Students will clearly want to avoid either of these outcomes, and thus will be motivated to work hard during the school year. The next section will discuss an innovative program recently implemented in Chicago that seeks to hold students accountable for their progress, yet gives failing students a second chance to exhibit their academic prowess.

The idea that testing would be more effective if something tangible were at stake is equally applicable to students and to their teachers and schools. Later on, I will discuss how school systems can use—and in some cases

already are using—students' test scores to identify schools that are failing, and some of the ways in which school administrations are intervening in such cases.

GRADUATION EXAMS AND EXIT EXAMS FOR SCHOOL LEAVERS

The above section argues in favor of regular monitoring of students as they progress through school. A closely related idea is to test students in grade 12 to ensure that they will graduate from high school with a skill level commensurate with the demands of the labor market. As noted in the next section, some states are moving in this direction and one—New York—has had a system similar to this one in place since the last century.

A less widely practiced policy is to provide a test of achievement that all students must pass before being allowed to graduate from high school *or to drop out of high school*. Such a test would ideally be offered to students in grade 9 or 10 and would focus on basic skills: reading, writing, and mathematics. In most states, such a policy would represent a sea change in how those likely to drop out of school are treated. In many states, students are allowed to drop out of school when they reach the age of sixteen, without having to demonstrate command of even the most basic skills. In today's economy—in which new technologies and changes in international trade patterns have acted to shift employers' needs toward more highly skilled labor—student dropouts have fared particularly badly over the last two decades. During the 1980s, the real wages of those with a college degree held fairly steady, while the real wages of those with a high school diploma or less fell substantially. For instance, Blackburn, Bloom, and Freeman (1990) report that between 1973 and 1987 the earnings of white male high school dropouts who worked full time and full year and were twenty-five to thirty-four years old fell from \$20,128 to just \$15,922. (Both of these figures are expressed in 1987 prices.) Such a precipitous fall in earnings over this short period suggests that schools should be particularly concerned about how well they are preparing their weakest students for the modern labor market. Since so many of these students ultimately drop out of high school, it stands to reason that the criterion for

dropping out should not simply be age, but should instead be a minimal level of achievement on a test of basic skills.

Another advantage of such a policy shift is that it gets the incentives right for students. A high school teacher will have little impact on a disaffected fifteen-year-old student who can feel free to ignore class assignments and so on, knowing that he or she can drop out at will after turning sixteen. By supplementing or replacing the policy of a minimum school-leaving *age* with a minimum school-leaving *level of achievement*, schools will give such students the right incentives to make the most of the time they do spend in school. Students will realize that if they cannot demonstrate mastery of a core set of skills—such as reading, writing, and basic mathematics—they will have to remain in school until they are able to do so. Of course, accommodation will be required for students with learning disabilities.

HIGHER GRADING STANDARDS

Another component of a school's overall standards is the way in which its teachers assign letter grades. If a school makes it overly easy to obtain a grade of A, the school is likely to reduce many students' effort. The best students, having obtained the top grade with little effort, will not find it worthwhile to work harder, because when they apply for a job or for admission to a university, their transcripts will not convey to potential employers this additional effort. For those who are not already A students, weaker grading standards can also be counterproductive in the sense that if any of these students think that a potential employer will care about letter grades, they can receive adequate letter grades without exerting much effort. Similarly, if B or C students hold any hope of attending a university, then letter grades should be of direct concern to them, so higher grading standards should induce additional effort.

HOMEWORK

The above suggestions deal with specific examples of how a school can set higher standards. Another step that a school could take is to set higher expectations of its students. A primary example of this is a school that encourages teachers

to assign more homework to students, especially in core subjects such as English and math. The next section provides a summary of a growing body of evidence supporting homework as a key to creating better schools.

THE EFFECTIVENESS OF HIGHER EDUCATIONAL STANDARDS IN PRACTICE

CURRICULUM STANDARDS AND TESTS OF ACHIEVEMENT: ARE THEY WIDELY USED? DO THEY WORK?

One of the main proponents of curriculum standards and achievement testing has been John Bishop. In a series of papers, Bishop (1996, 1997) has gathered indirect evidence that such standards can palpably improve the quality of education. For instance, he reports that students from the state of New York tend to outperform students from other states on standardized tests such as the Scholastic Aptitude Test (now known as the SAT). While there are literally dozens of potential explanations for such a finding, one is that New York has stood alone in setting a statewide exam for high school seniors: students in New York have long had the option of taking the Regents examinations. Those who pass the exams receive a high school diploma different from that received by students who opt not to take the Regents exams. Passing the exams has in general been a prerequisite for college entry in New York.

New York is now in the middle of a quite bold experiment, in which the alternative high school diploma, known as the local diploma, is being phased out. Current plans are for the Regents exams to become a requirement for high school graduation for all students in the state by 2003 (*New York Times* 1996).

Bishop (1996, 1997) has also analyzed the performance of Canadian students on the 1991 International Assessment of Educational Progress (IAEP). Canada's educational system is in many ways similar to that of the United States, but significantly, many of Canada's ten provinces now require students to pass a provincewide exam before graduating from high school. Bishop finds that students from the provinces that have implemented graduation exams tend to perform significantly better on

the IAEP. A clear concern in all empirical research that uses differences in policies across political borders is that variations in the given policy might be endogenous. That problem appears especially likely in this context: for political reasons, a ministry of education is less likely to institute provincewide testing if it knows that its students are likely to fare poorly. Since research consistently finds that the socioeconomic background of parents is highly predictive of students' achievement, it could be that only richer provinces would institute testing in the first place. However, the provinces that have instituted testing include not only the most economically developed, but also the least economically developed: Newfoundland. This suggests that the reported correlation between the existence of graduation exams and student test scores reflects true causation. Bishop (1996) also reports corroborating facts. For instance, he finds that in provinces that have instituted graduation exams, students report watching less television.

In the United States, how widespread is the idea of graduation exams across states? New York has offered the Regents examinations to high school students for over a century. But until 2003, these exams will be partly voluntary; students can instead opt for the local diplomas that do not require the test. In many other states, work is under way to develop curriculum standards, and in a subset of these states plans are also under way to require high school seniors to write graduation exams that are linked to the curriculum.

The American Federation of Teachers (AFT) has recently committed to performing an annual evaluation of states' progress in developing precisely worded curriculum standards and tests. Their 1996 report suggests a crazy quilt of reforms across the states. The AFT finds that all states apart from Rhode Island and Wyoming are now developing grade-by-grade content standards. Unfortunately, there is an extraordinary disparity in the level of detail provided in each of the state curricula. The AFT reports that only Virginia has developed curriculum standards in English, math, science, and social studies that are sufficiently clear and explicit to provide guidance to teachers and parents about what should be taught in each grade. Of the forty-eight states that are writing or have

written curriculum standards in at least one of these subjects, forty-two have developed or are developing tests that will be based on the content of the curriculum. However, only fifteen of these forty-two states have planned or are planning test programs in all four subject areas mentioned above, with tests being based on precisely worded content standards. In summary, most states are now developing content standards in at least a few of the key subject areas and are developing statewide tests of these curricula. But only a minority has as yet developed a comprehensive set of content standards backed by testing.

ACHIEVEMENT TESTS, REMEDIATION, AND GRADE RETENTION POLICIES

Earlier, I argued that a specifically worded curriculum backed by periodic testing of students could by itself significantly improve school quality, because it provides an objective report card on individual student achievement. When a student falls behind grade level, the student, his or her parents, and the school's teachers can react. But it seems reasonable that content standards and testing will be more effective if school systems have a formal plan in place to deal with students falling behind grade level. Another critical question is whether the student has anything at stake. Will a student who is far behind grade level have to attend after-school tutorials, summer school, or—in extreme cases—even repeat the same grade next year? Alternatively, do students not take the tests seriously, because there are no consequences attached to poor performance on them? And if the school system does mandate remediation, has it put in place a funding mechanism?

The AFT (1996) report provides partial answers to these questions. Its survey reveals that eighteen of the fifty-one states (the District of Columbia is treated as a state) have state-mandated remediation for students who do not meet the state's educational standards. However, only ten of these eighteen states also provide funding to schools for the additional teaching.

What are the state-mandated consequences for students who fail to do well on the achievement tests? The AFT study finds that even though forty-two states have planned or are planning statewide tests, only a handful

have made promotion between grades or graduation from high school conditional upon test scores. The more common action, adopted or about to be adopted by thirteen states, has been to institute graduation exams that test whether a grade 12 student achieves at a level equivalent to grade 10 standards or higher. This is a noteworthy trend.

However, it is surprising how few states have made student promotion decisions based in any way on students' results on the state tests. The only states or districts to have done so by 1996 were the District of Columbia, North Carolina, and South Carolina.

Why have only three of the forty-two states with tests and curriculum standards linked test performance to grade promotion? One explanation is that the empirical literature on the consequences of grade retention has in general found that holding a student back one year has either zero effect or a negative effect on the student's subsequent rate of learning. For instance, in a review of the literature, Holmes (1989) concludes that most studies have found that grade retention is associated with poorer performance after the student is held back a year. Only nine of sixty-three studies found that retention improved student performance. Holmes indicates that in most of these positive studies the "treatment" of students was not simply retention but retention accompanied by quite intensive remediation. It is therefore not clear whether students who were held back a year did better than the comparison group because of the additional year or the extra help they received.

The above summary of state policies on curriculum standards and testing suggests that most states are now working toward these goals, but that in most cases much remains to be done. It also highlights the substantial diversity across states in their policies concerning educational standards.

The state-level summary ignores the fact that in most states individual school districts enjoy considerable autonomy to create their own programs to supplement or strengthen statewide initiatives. For this reason, the above summary is likely to understate the extent to which students in public schools are held to curriculum-based standards. Similarly, the summary is also likely to understate the disparities across schools in the stringency of

academic standards. Interschool and interdistrict variations in curriculum content and in standards are of particular concern given that each year so many children switch schools when their parents move. For this reason, it would seem worthwhile for a state to set high standards in order to level out any existing variations in standards across districts.

A detailed analysis of how individual school districts within even one state set and enforce standards would be a major undertaking. But certain school districts around the country have received national attention for their innovations in setting academic standards. The next section discusses a particularly bold set of reforms that the Chicago public schools have recently implemented.

THE CHICAGO PUBLIC SCHOOLS EXPERIMENT

The Chicago public schools (CPS) system is one of the largest school districts in the country. Its students represent an ethnically and racially diverse group that must contend with all the challenges of life in a modern urban area. It is perhaps not surprising, then, that in Chicago students' academic performance has long lagged behind national averages. For example, throughout the 1990s grade 9 students' average performance on the Tests of Achievement and Proficiency (TAP) has hovered between the twenty-third and thirty-sixth percentiles of national norms. Results for other grades have fallen into a similar range.⁴

The CPS has given tests to children in various grades throughout the 1980s and 1990s. But during the 1996-97 school year, school administrators made the tests "matter." In grades 3, 6, 8, and 9, students whose performance lagged behind national norms on either the reading or mathematics portion of the tests were required to attend summer school. The cutoff points below which students were required to attend summer school were 2.8 for grade 3, 5.2 for grade 6, 6.8 for grade 8, and 7.9 for grade 9. (The tests were given in the spring, so that a student progressing at the normal rate would be at grade level at the time of the test, while a score of 2.8 in grade 3 would suggest that the student was approximately two months behind national norms.) The summer school, known as the Summer Bridge Program, lasted for six to seven weeks. At the end of the Summer Bridge Program, students took the tests a second

time—the TAP in grades 8 and 9 or the Iowa Tests of Basic Skills in lower grades. A student who met the cutoff grade equivalents listed above for both reading and math by the end of summer was allowed to advance to the next grade.⁵ Students who did not reach the cutoff level in either test were required to repeat the grade.⁶ During the 1997-98 school year, students who were held back were in many cases to receive additional help, typically in the form of tutorial classes.

Results from the 1996-97 school year are quite remarkable, both for the sheer number of students who failed the initial tests in spring, and for the sizable gains in achievement recorded for those who entered the Summer Bridge Program. Table 1 reports the number of students who failed to reach the cutoff in either math or reading in

Table 1
RESULTS FROM THE CHICAGO PUBLIC SCHOOLS' TESTING AND SUMMER BRIDGE PROGRAMS

Variable	Grade 3	Grade 6	Grade 8	Grade 9
SPRING TEST RESULTS				
Students taking math test	23,989	25,275	22,708	22,986
Students taking reading test	24,124	25,311	22,776	22,967
Students taking at least one test	24,124	25,311	22,776	22,986
Students failing at least one spring test	11,632	8,870	6,180	14,287
As a percentage of students taking at least one spring test	48.2	35.0	27.1	62.2
SUMMER PROGRAM AND SUMMER TEST RESULTS				
Students taking summer test	10,336	8,275	5,831	9,610
As a percentage of students who should have taken summer test	88.9	93.3	94.4	67.3
Students passing summer test	4,236	3,668	2,891	3,695
As a percentage of students taking summer test	41.0	44.3	49.6	38.4
As a percentage of students who should have taken summer test	36.4	41.4	46.8	25.9
Percentage of all students taking spring tests who passed by end of summer	69.3	79.4	85.6	53.9
GAINS IN PERFORMANCE AMONG SUMMER STUDENTS				
MEAN GRADE EQUIVALENT (GE) OF STUDENTS				
Spring GE of students below promotion level on spring reading test	2.03	4.41	5.88	6.32
August GE of students below promotion level on spring reading test	2.47	5.11	6.87	7.79
Gain in reading GE over summer	0.44	0.70	0.99	1.47
Spring GE of students below promotion level on spring math test	2.29	4.58	6.25	6.69
August GE of students below promotion level on spring math test	2.98	5.20	7.04	8.03
Gain in math GE over summer	0.69	0.62	0.79	1.34

the spring tests, both as a raw number and as a proportion of all CPS students who took the spring tests at that grade level.⁷ It shows that 27.1 to 62.2 percent of students failed at least one of the two tests, depending on the grade level.

The table also reports outcomes of the Summer Bridge Program. There are two complications that must be dealt with. First, I exclude from the analysis students who had met the cutoff scores in the spring test but who enrolled in the Summer Bridge Program on the advice of the school because they were near the margin. Such students automatically “pass” the test in August. The second complication derives from the fact that only a fraction of the students who should have enrolled in the Summer Bridge Program took the tests at the end of summer. For this reason, I calculate pass rates at the end of summer using in the denominator both the total number of students who took the summer test and the total number who *should* have taken the summer test. Using the latter number, I find that 25.9 to 46.8 percent of students who should have taken the summer test passed. Note that a substantial fraction of students in grade 9 did not take the summer test. The most likely explanation for this is that by CPS policy, students who were fifteen by December 1, 1997, were not to be held back in grade 9. When the pass rates are calculated as a percentage of those who actually took the summer tests, the success rate is much higher, ranging from 38.4 to 49.6 percent, with the highest success rate observed among grade 8 students. But overall, well over half of the regular students in these grades are promoted at the end of the school year. The highest success rate is 85.6 percent, among grade 8 students.

Clearly, the Summer Bridge Program is not a panacea for students who initially obtain low scores: over half of those enrolling in the program do not pass in their second attempt. But on average, students progressed remarkably during the program. Table 1 also shows the initial mean grade equivalent of those who failed to meet the promotion criterion in the given subject, and the mean grade equivalent that these same students obtained after the Summer Bridge Program. The mean gain in grade equivalents is typically one half year to a full year or even

more. Given the low base from which these students began in the spring, they remain on average one-half grade to one grade equivalent behind by August, or in the cases of reading scores in grades 8 and 9, slightly more than one grade equivalent behind. The observed improvements, however, are extremely impressive.

Of course, one concern raised by these data is that we are observing “regression to the mean.” If a student has a bad day when taking the spring test and scores considerably below his or her potential, that student might have to enroll in the Summer Bridge Program. In such a situation, the student might score much higher on the second test in August, not because of the usefulness of the program but because the low test score in the spring was a statistical aberration. It would be useful to study patterns in the test scores of students *prior* to the year they are required to attend summer school, to determine whether some of the summer school students were forced to attend simply because of a bad day in the spring test. Similarly, it would be useful to follow these students for at least one more year to check whether the remarkable gains in achievement over the summer endure.

Nevertheless, the fact remains that nearly half of Chicago’s summer school students had by the end of the summer improved sufficiently to meet the promotion criterion. Improvement on this scale suggests that the Chicago public schools system has found an ideal incentive system for students. The impressive gains in mean grade equivalent, of anywhere from 0.6 to 1.3 years in the space of the six- to seven-week Summer Bridge Program, point in the same direction. Low academic achievement now has tangible consequences for students: the students must spend extra time in remedial classes, both in the summer and during the school year. The CPS has enforced reasonably high standards and gotten the incentives right at the same time that it has directed additional funding toward students whose achievement has lagged the most. Over time, this fledgling program could do much to improve the academic achievement of students who are most in need.

The CPS program also improves on the traditional “fix” for education, in which wholesale increases are made in school spending. By identifying students most in need

through tests each spring, scarce financial resources are being targeted toward students who truly need additional time with teachers.

It is possible to compare at least roughly the costs and benefits of regular schooling with those of the summer school program. A typical American school spends about \$5,000 per pupil per year, and the typical student in that school will gain one grade equivalent over the school year. In 1997, the CPS spent about \$34 million on its Summer Bridge Program, or about \$720 per participant. This sum translates into about \$1,000 for each student who took the test in August, since some marginal students who participated in the Summer Bridge Program voluntarily did not take the test at the end of summer. This latter figure represents about one-fifth the cost of a year of regular schooling for each student who participated. Yet the gain in achievement far surpasses one-fifth of a grade equivalent. As shown in Table 1, the mean gain over the six- to seven-week summer program ranged from a low of 0.44 of a grade equivalent in grade 3 reading tests to 1.47 grade equivalents in grade 9 reading tests. Even assuming that these students gained a full grade equivalent during the regular academic year, on a dollar-for-dollar basis the Summer Bridge Program is anywhere from 2.2 times to 7.4 times as effective as schooling during the regular year. By any standard, these additional expenditures appear to have been much more effective than traditional expenditures made during the academic year. Over time, it will become possible to test whether these gains are permanent or transitory.

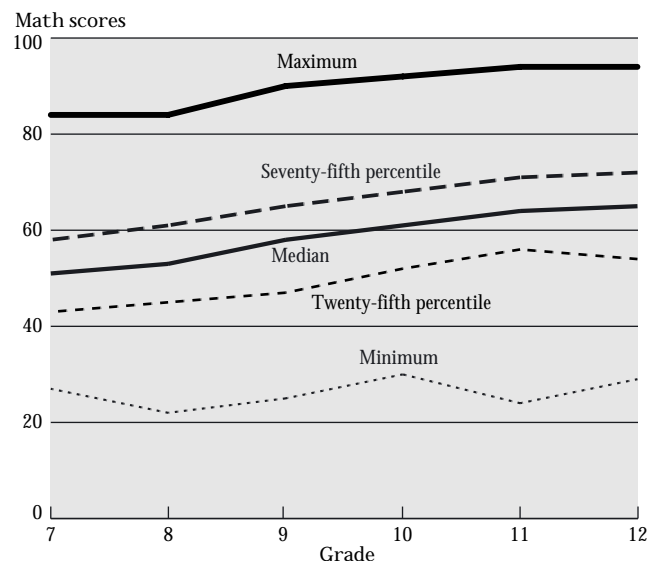
Some readers will rightly wonder whether the initial failure of a number of students to meet the promotion criteria simply reflects adverse conditions in the Chicago public schools relative to those in public schools elsewhere. Consider the following simulation. Suppose that national standards were put in place dictating that any secondary school student more than a year behind in mathematics achievement was required to attend summer school. Suppose that the criterion for being “more than a year behind” was that a student in one grade had a test score below the median test score of students in the previous grade. What would happen?

Chart 4 depicts the median math test score in a representative sample of American high school students by grade level, which I calculated from the Longitudinal Study of American Youth (LSAY). (This study followed approximately 6,000 students over a five-year period.) The chart also shows the test scores of students at the twenty-fifth and seventy-fifth percentiles, and the minimum and maximum scores observed in each grade. The most striking aspect of the chart is the huge disparity in achievement among students within any grade level. In the sample, the increase in the median math test score between grades is approximately 2.8 points out of 100. This median gain is tiny compared with the dispersion within any grade: the gap between the test scores of students in the twenty-fifth percentile and the scores of students in the seventy-fifth percentile within a grade is typically about 16 points.

The slow growth in student performance *across* grades relative to the large degree of dispersion *within* grades in this nationally representative data set suggests that Chicago’s experience is not atypical. The LSAY data clearly indicate that large proportions of students lag

Chart 4

Math Test Scores by Grade in the Longitudinal Study of American Youth



Source: Author's calculations, based on data from the Longitudinal Study of American Youth.

Table 2
**PERCENTAGE OF STUDENTS WHOSE MATH SCORES
 WOULD MISS CUTOFFS BASED ON MEDIAN SCORES
 IN PRIOR GRADES**

Grade	Percentage below Median Score in Previous Grade	Percentage below Median Score Two Grades Earlier
8	40.8	N/A
9	36.7	31.6
10	39.0	26.4
11	39.1	29.2
12	45.5	39.7

Source: Author's calculations, based on math test score data for students participating in the Longitudinal Study of American Youth.

behind national norms in schools across the country. Table 2 shows the percentage of students in the LSAY who would have to repeat a grade if the criterion for promotion was that a student in grade “N” needed a test score equal to or above the median score among students in grade “N-1.” The percentages of students who would be required to repeat each grade are extremely high, and quite similar to what has been reported in Chicago, where the promotion criterion is that students must score, roughly speaking, within a year of the national norm for their grade level. The table also shows the percentage of students who would have to attend summer school under lower standards. Even when the promotion criterion is reduced so that a student needs only to obtain the median test score of students *two* grades back, anywhere from 26 percent of students in grade 10 to 40 percent of students in grade 12 would miss the cutoff point.

This simulation illustrates the extent of the problem in American public schools: disparity in achievement within grades is simply huge. It also shows that when school districts set reasonably high standards, they should be prepared for a large proportion of their students to fall below the standard, at least initially.

GRADING STANDARDS

A little-examined characteristic of schools is the stringency with which they grade students. If a school makes it easier to obtain a letter grade of A or B, students might respond by exerting less effort. Such a response will occur if students care directly about letter grades, as opposed to their true level of academic achievement. There are several

plausible reasons for this. First, students may care about letter grades simply because their parents care. Second, students may realize that employers may use letter grades as a signal of a student’s achievement, willingness to work hard, and so on. Some employers may examine high school transcripts directly, but it seems likely that an indirect mechanism is at least as important: firms rely on the postsecondary sector to identify high school students who have done well. College admission offices in effect act as an information intermediary between students and employers by closely examining high school transcripts in an effort to identify the students most likely to gain from further education.

In Betts (1997), I examine grading policies in math and science courses in a representative sample of American schools. By comparing students’ test scores in these subjects with their letter grades, I construct measures of the grading standards at each school. I find that American high schools differ radically in the way in which they assign letter grades to students of given achievement. I also find that the stringency with which a school grades is strongly related to the rate at which students learn. Even after controlling for the initial level of achievement of students at the school, traits of the individual student, traits of his or her family and peers, and detailed traits of the classroom, I observe that the school’s grading standard remains a highly significant and positive predictor of gains in test scores. Unfortunately, a policy of higher standards does not improve the performance of all students identically. Although C students benefit from attending a more rigorous school, A students benefit even more. The lesson is clear: in this instance, a policy of higher standards will also induce higher disparities in achievement. If a school administration is concerned with the distribution of student achievement, and not just the mean level of achievement, remedial policies to assist students with lower grade-point averages are in order.

HOMEWORK

Each of the above suggestions for improving schools involves setting higher standards. A closely related suggestion is to increase the amount of homework that

teachers assign. This is perhaps better thought of as a way of setting higher academic expectations than as a way of setting higher academic standards, although of course teachers can hold students accountable if they routinely fail to complete their homework. A policy of assigning more homework is quite similar in spirit to the earlier suggestions, because it too recognizes that one of the most important inputs in the “education production function” is the student’s own effort. This simple fact has been ignored in most of the traditional studies of whether “spending matters.”

A number of studies of the impact of homework on achievement have been carried out. Cooper (1989) gives a good review of the existing evidence. Two experimental studies have been performed, with somewhat mixed results. In both cases, the number of students involved in the experiment was very small, ranging from roughly 90 to 350 students. Cooper also reports that a number of correlational studies find a positive and significant link between the time students spend on homework and their achievement. The effectiveness of homework appears to be higher in secondary schools than in the elementary grades. Unfortunately, virtually all of the correlational studies test for a relationship between the amount of homework that students report doing and their achievement. This approach leads to a clear possibility of reverse causation: if better students routinely *choose* to do more homework than their peers, then the observed relationship between test scores and homework might be spurious.

A partial solution to this problem is to model student achievement not as a function of the homework that students do but as a function of the homework that teachers assign. In Betts (1996b), I use the aforementioned LSAY data for this purpose. I find a strong positive link between the amount of homework that teachers assign and the rate at which the student’s test score rises. I also estimate a “fixed-effect” model in which I use *variations* in the amount of homework assigned to individual students across grades to identify the effectiveness of homework, with similar results. The estimated effects of additional homework are quite striking. Using published estimates of the relationship between math test scores and

earnings, I calculate that an hour spent doing homework is equivalent to earning about \$6 to \$12. I make these estimates by discounting the future wage gains at a rate of 3 percent and by setting the opportunity cost of doing homework at the average earnings per hour of teenagers. Additional homework appears to be a particularly cost-effective method of improving school quality: it is the total amount of homework assigned—rather than the amount assigned, graded, and returned—that is more closely related to students’ rate of learning. The effectiveness of homework, by any measure, is quite large. The results suggest that among the students in grades 7 through 12 in the sample, a one-hour homework assignment is as effective as an hour spent in class.

HIGHER STANDARDS ARE NOT ONLY FOR STUDENTS: PROVIDING INCENTIVES TO TEACHERS AND SCHOOL ADMINISTRATORS

To this point, I have focused on the methods of improving incentives for students that target additional spending toward the students most in need. However, the same principle of tying additional spending to the setting of higher standards can be applied to teachers and entire school systems just as easily as it can be applied to students. Space constraints prevent me from developing this theme in detail, but the following discussion highlights the main arguments.

The essential point is that there are good teachers and bad teachers, effective principals and less effective principals. What, then, should a school board do when, after setting higher standards for its students, it realizes that at some schools the only remedy for low achievement is to improve the quality of teaching?

Improving teacher quality requires a two-pronged approach—setting up the economic incentives required to attract well-trained college graduates to the teaching profession while providing opportunities for more experienced teachers to gain new skills. Typically, school districts set teachers’ salaries as a fairly rigid function of teachers’ years of experience and the degrees they hold. To obtain a permanent certification, teachers in many states must obtain a certain number of graduate-level credit hours, which

typically lead to a Master's degree. Yet the evidence that teachers who hold a Master's degree are better teachers is decidedly mixed (see, for instance, Betts [1995a], Grogger [1996], and Betts [1996a]). Paying teachers by the level of degree held may make less sense than paying teachers extra for any college courses that pertain to their field of teaching in the school. A number of authors—for instance, Goldhaber and Brewer (1997) and Betts (1995b)—have found that college training in the field taught is related to teaching ability. Similarly, mechanically linking teachers' pay to years taught may not be the optimal policy: evidence suggests that teachers, especially after the first few years of teaching, do not necessarily continue to improve their quality of teaching much over time. Similarly, it is important for school administrators to respond to market forces. Murnane et al. (1991) show that over the last twenty years the starting salary for teachers has consistently lagged behind that of college graduates who work in industry. They argue strongly in favor of merit pay for teachers as a method of ensuring that the best teachers remain in the profession.⁸

The question immediately arises, how can principals identify the best teachers in order to allocate merit pay? The need for ongoing and objective assessment of teachers provides yet another reason why it is so important for school districts to set out a clear curriculum and then to assess students. But it would be wrong to reward teachers solely on the basis of the mean test scores in their classes. Given evidence that family background and peer effects strongly influence student achievement, such a policy would in effect punish teachers who taught in disadvantaged neighborhoods. Within schools, it would aggravate any tendency that might already exist for teachers to prefer to teach the most advanced students within each grade. A more reasonable approach might be to establish merit pay as a function of how much student achievement *improves* over one or two years.

Similar policies of merit pay for principals might also work. A number of school districts around the country have gone further, setting the pay of superintendents of entire districts to reflect the rate of improvement of the district's students.

Merit pay, additional course work for experienced teachers, and a less rigid structure for teacher salaries that would allow schools to attract talented college graduates in fields such as math and science are all good ideas. In some cases, however, more radical solutions might be needed. Perhaps in recognition of this, a number of state legislatures over the last few years have attempted to remove teacher tenure, so that school districts would find it easier to fire teachers who were not performing adequately. To the best of my knowledge, none of these reforms has met with success, because of opposition from a number of sources, including—not surprisingly—teachers' unions. Individual school districts have also started to put on "probation" those schools whose students fare poorly, and in some cases have "reconstituted" entire schools by firing or reassigning virtually all employees, from the principal down to the custodial staff, in a bid to change the prevailing culture at the school. It is too early to know whether such radical restructuring has had the intended results.

A different method for improving schools is to increase the degree of competition between schools. Basic economic models argue that when a firm has a monopoly, it will restrict supply and charge higher prices than it would if there were a high degree of competition in the market. By analogy, when a school district is the sole provider of schooling, the lack of competition allows it to do less with each education dollar than it would under competition. Ballou (1996) provides a recent and interesting example of how a lack of competition may render public schools inefficient. He finds that when hiring teachers, public schools do not seem to give any preference to applicants who have superior academic records. This finding is puzzling, given evidence by Ehrenberg, Goldhaber, and Brewer (1995) that cognitive abilities of teachers are positively and significantly related to the rate at which their students learn. Ballou's conclusion is that public schools face little competition for students, and so do not invest sufficient effort in finding the best applicants for teaching jobs. This conclusion, if true, would be an example of an inefficiency that would surely disappear if schools competed with each other more strongly for students.

A companion paper in this volume by Hoxby (1998) describes in detail the evidence that competition—whether between school districts or between public and private schools—might improve the efficiency of public schools. In addition, the paper by Rouse (1998) in this volume addresses the effectiveness of vouchers.

While the argument in favor of increased school choice appears to have some empirical backing, it is crucial to understand that such choice can work only if parents make informed decisions. Informed decisions require good information about schools. This provides yet another rationale for regular testing of students. Without a districtwide—or, preferably, city- or statewide—report card on student achievement and gains in student achievement at each school, it is unlikely that parents will be able to make informed decisions about the schools that are best for their children.⁹

OBSTACLES TO HIGHER STANDARDS AND SOME SUGGESTED SOLUTIONS

If academic standards—in the form of a clearly worded curriculum and tests designed around the curriculum—are such good ideas, then why do we not see more widespread use of these tools? Critics have raised many specific objections, but it is crucial to realize at the outset that testing threatens many interests. What politician wants to have it publicized that schools within his or her district are not adequately serving students? Testing can also cause discomfort for teachers, students, and in some cases parents. Because assessment, done properly, provides an objective “report card” on schools, it often provides impetus for radical change. Many professionals in education will naturally resist change.

Existing theoretical work on educational standards points to a second reason for opposition to higher standards. Two models developed by Costrell (1994) and Betts (1998) differ in some regards, but both establish that whenever a school raises its standards (such as a pass-fail standard), some students will lose out. In essence, a student whose ability or diligence was such that he or she was initially indifferent to meeting the standard or falling below it will choose not to exert the extra effort required after the standard is raised. This can lead to a significant

drop in well-being for such students. This theoretical result suggests that in the real world, school administrators who raise standards will have to devote additional attention to “borderline” students to ensure that they are able to continue meeting the requirements for grade promotion or for high school graduation.

A commonly heard complaint about testing of curriculum standards is that it will lead to teachers wasting time “teaching to the test.” Why should teachers be forced to squander valuable class time helping children to memorize facts and tricks for a multiple-choice test when they should be steering children toward more profound forms of learning, or so the argument goes. There is no doubt an element of truth in this statement. But, ideally, a test should not only require regurgitation of memorized facts, but should also assess a student’s ability to synthesize, to apply concepts learned in one context in a new environment, and so on. Ideally, then, tests will include not only multiple-choice questions but also questions requiring a written response. In short, if administrators write a test properly, teaching to the test is exactly what teachers should be doing.

Recent experience indicates that the main barriers to higher educational standards backed by testing are political. President Clinton’s call for voluntary national tests in reading in grade 4 and in math in grade 8 has recently met strong opposition on Capitol Hill.¹⁰ Some legislators have objected that federally backed tests represent an intrusion by Washington into education, which traditionally has been controlled at the state level. President Clinton has responded, with some justification, that the basic elements of mathematics are the same regardless of the state in which the student in question lives. Why should children—and schools—in some states be held to standards lower than those in other states?

One can also make a case for national standards on cost grounds. Surely, a reasonable set of national standards could be drafted at far lower expense than could fifty sets of standards, each specific to a given state. The problem is compounded by the actions of the many school districts that have drafted their own content standards in recent years. Not only does this lead to needless duplication of

effort, but it can create problems for students who move between school districts within a state at some time.

Clearly, however, when representatives from fifty states meet to attempt to establish national standards, diverse opinions are likely to lead to diluted national standards. Fear of such an outcome may explain why so many states have taken it upon themselves to develop curriculum standards, and why many school districts have developed their own content standards to supplement those provided by the state. It would appear that proponents of national standards will have the greatest chance of success if they focus on subjects in which there is general agreement about required elements of a core curriculum. For instance, the relative success of the National Council of Teachers of Mathematics in its attempt to create national standards in math stands in strong contrast to the recent failure of an attempt to create national standards in the much more contentious subject area of history. The divergence in outcomes may reflect underlying disagreements about what is important in an area. Therefore, it might be advisable, at least at first, for educators seeking national or even state-level standards to focus on areas such as mathematics, reading, and writing.

Space constraints prevent a further discussion of barriers to higher standards. However, readers interested in these issues should see Ravitch (1995) for a compelling insider's account of the history of educational standards in the United States. In particular, Chapters 5 and 6 of her book develop a clear prescription for change that is appealing in an economic sense, yet takes heed of the political realities of school reform today.

CONCLUSION: A CHECKLIST FOR REFORM

In the 1990s, virtually all states have started to develop curriculum standards to increase student achievement. Although the states vary remarkably in the number of subject areas for which they are developing curriculum standards—and in the specific content of the standards—as a group they have clearly made significant progress.

A necessary companion policy to higher standards is increased spending on assessment of student achieve-

ment. Testing is crucial. To ensure that reforms are effective, administrators must do more than prescribe a minimal curriculum in key subjects; they must evaluate the extent to which individual students meet the standards. Such tests serve a dual purpose: not only do they create an incentive for students to exert effort, they also provide a means to make teachers, schools, and entire school systems accountable to the public. The regular publication of test results by school is also an indispensable tool if public schools introduce a system in which parents can choose the school their children attend.

In this paper, I have reviewed recent attempts by states and school boards to raise standards. Limited empirical evidence suggests that higher expectations—whether established through higher graduation standards, more stringent homework requirements, higher grading standards, or increased requirements for promotion between grades—can spur student achievement. The paper also reviews the serious roadblocks that have hindered attempts to tighten standards, and suggests some ways in which school administrators can address the concerns of critics of testing and standards.

The two most historically important reforms to public schools in this century—raising the school-leaving age and increasing spending per pupil—in a sense form a two-legged stool. Allowing students to leave school at a certain age without having them demonstrate a minimal level of achievement is a shortsighted policy. Achievement, not age alone, should determine when a person is ready to leave school. Similarly, large increases in school spending that are not accompanied by increases in the standards to be met by students and their schools are likely to achieve little. Only by coupling these two policies to higher standards—and by testing the ability of students to meet the standards—are we likely to see large improvements in school quality. For this reason, the reforms in educational standards currently under way in many states have the potential to be surprisingly more effective than previous reforms.

ENDNOTES

The author thanks Sandra Storey and Joseph Hahn for helpful discussions concerning the Chicago data.

1. Male workers are chosen, since the studies reviewed in Betts (1996a) examine earnings of men only.
2. In 1990-91, total spending per pupil in American public schools was \$5,320, while the current expenditure was \$4,847 per pupil (National Center for Education Statistics 1991, p. 155).
3. The one exception is for estimates of the return to increasing the teacher-pupil ratio derived from school-level studies. Betts (1996a) finds that the average effect of this intervention is actually slightly negative, so that increasing the discount rate lowers the predicted losses.
4. All information in this section that relates to the Chicago public schools was obtained directly from the CPS system, except where noted.
5. A student who passed one test in the spring and the other test at the end of summer school was promoted. In addition, a small number of students who met the grade equivalent criterion for both reading and mathematics in the spring, but whose scores were borderline, voluntarily enrolled in the Summer Bridge Program.
6. The consequences for grade 9 students whose test scores were too low were slightly different. Any such student who was at least fifteen years old by December 1, 1997, was not retained, but was sent to a special remedial school, known as a High School Transition Center.
7. Students enrolled in the bilingual education program are exempted from participation in the Summer Bridge Program for up to three years. Similarly, special education students are not required to meet the standards. Accordingly, all calculations in Table 1 exclude these two types of students.
8. See, however, Chapter 5 of Ballou and Podgursky (1997), which argues that attempts to strengthen the incentive structure faced by teachers are unlikely to have much success, given the likelihood of opposition from entrenched interests.
9. See Hanushek et al. (1994, Chap. 6) for a detailed summary of ideas on how administrators could use incentives to improve schools.
10. See, for instance, Applebome (1997).

REFERENCES

- American Federation of Teachers*. 1996. MAKING STANDARDS MATTER, 1996: AN ANNUAL FIFTY-STATE REPORT ON EFFORTS TO RAISE ACADEMIC STANDARDS. Washington, D.C.
- Applebome, Peter*. 1997. "President's Proposal for National Reading and Math Tests Is Facing Wide Opposition." NEW YORK TIMES, August 3: 14.
- Ashenfelter, Orley, and Alan B. Krueger*. 1994. "Estimates of the Economic Return to Schooling from a New Sample of Twins." AMERICAN ECONOMIC REVIEW 84: 1157-73.
- Ballou, Dale*. 1996. "Do Public Schools Hire the Best Applicants?" QUARTERLY JOURNAL OF ECONOMICS 111: 97-133.
- Ballou, Dale, and Michael Podgursky*. 1997. TEACHER PAY AND TEACHER QUALITY. Kalamazoo, Mich.: W.E. Upjohn Institute for Employment Research.
- Betts, Julian R.* 1995a. "Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth." REVIEW OF ECONOMICS AND STATISTICS 77: 231-50.
- _____. 1995b. "Which Types of Public School Spending Are Most Effective? New Evidence on the School Quality Debate." University of California, San Diego, Department of Economics Discussion Paper no. 95-03.
- _____. 1996a. "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature." In Gary Burtless, ed., DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS, 141-91. Washington, D.C.: Brookings Institution.
- _____. 1996b. "The Role of Homework in Improving School Quality." University of California, San Diego, Department of Economics Discussion Paper no. 96-16.
- _____. 1997. "Do Grading Standards Affect the Incentive to Learn?" University of California, San Diego, Department of Economics Discussion Paper no. 97-22.
- _____. 1998. "The Impact of Educational Standards on the Level and Distribution of Earnings." AMERICAN ECONOMIC REVIEW 88, March.
- Betts, Julian R., and Jamie L. Shkolnik*. Forthcoming. "The Effects of Ability Grouping on Student Math Achievement and Resource Allocation in Secondary Schools." ECONOMICS OF EDUCATION REVIEW.
- Bishop, John H.* 1996. "The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning." INTERNATIONAL JOURNAL OF EDUCATION RESEARCH 23: 653-752.
- _____. 1997. "The Effect of National Standards and Curriculum-Based Exams on Achievement." AMERICAN ECONOMIC REVIEW 87: 260-4.
- Blackburn, M. L., D. E. Bloom, and R. B. Freeman*. 1990. "The Declining Economic Position of Less Skilled American Males." In Gary Burtless, ed., A FUTURE OF LOUSY JOBS? THE CHANGING STRUCTURE OF U.S. WAGES, 31-67. Washington, D.C.: Brookings Institution.
- Cooper, Harris*. 1989. HOMEWORK. New York: Longman.
- Costrell, Robert M.* 1994. "A Simple Model of Educational Standards." AMERICAN ECONOMIC REVIEW 84: 956-71.
- Ehrenberg, Ronald G., Daniel D. Goldhaber, and Dominic J. Brewer*. 1995. "Do Teachers' Race, Gender and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988." INDUSTRIAL AND LABOR RELATIONS REVIEW 48: 547-61.
- Goldhaber, Daniel D., and Dominic J. Brewer*. 1997. "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity." JOURNAL OF HUMAN RESOURCES 32: 505-23.
- Grogger, Jeff*. 1996. "Does School Quality Explain the Recent Black/White Wage Trend?" JOURNAL OF LABOR ECONOMICS 14: 231-53.
- Hanushek, Eric A.* 1996. "School Resources and Student Performance." In Gary Burtless, ed., DOES MONEY MATTER? THE EFFECT OF SCHOOL RESOURCES ON STUDENT ACHIEVEMENT AND ADULT SUCCESS, 43-73. Washington, D.C.: Brookings Institution.
- Hanushek, Eric A., et al.* 1994. MAKING SCHOOLS WORK: IMPROVING PERFORMANCE AND CONTROLLING COSTS. Washington, D.C.: Brookings Institution.
- Holmes, C. Thomas*. 1989. "Grade Level Retention Effects: A Meta-Analysis of Research Studies." In Lorrie A. Shepard and Mary Lee Smith, eds., FLUNKING GRADES: RESEARCH AND POLICIES ON RETENTION, 16-33. London: Falmer Press.
- Hoxby, Caroline M.* 1998. "What Do America's 'Traditional' Forms of School Choice Teach Us about School Choice Reforms?" Federal Reserve Bank of New York ECONOMIC POLICY REVIEW 4, no. 1.

REFERENCES (*Continued*)

- Lang, Kevin, and David Kropp.* 1986. "Human Capital Versus Sorting: The Effects of Compulsory Attendance Laws." *QUARTERLY JOURNAL OF ECONOMICS* 101: 609-24.
- Mincer, Jacob.* 1974. *SCHOOLING, EXPERIENCE, AND EARNINGS.* New York: Columbia University Press.
- Murnane, Richard J., Judith D. Singer, John B. Willett, James J. Kemple, and Randall J. Olsen.* 1991. *WHO WILL TEACH? POLICIES THAT MATTER.* Cambridge: Harvard University Press.
- National Center for Education Statistics.* 1991. *DIGEST OF EDUCATION STATISTICS, 1991.* Washington, D.C.: U.S. Government Printing Office.
- New York Times.* 1996. "Passing of Regents Exams to Be Required." April 24: A1, B9.
- Psacharopoulos, George.* 1985. "Returns to Education: A Further International Update and Implications." *JOURNAL OF HUMAN RESOURCES* 20: 583-604.
- Ravitch, Diane.* 1995. *NATIONAL STANDARDS IN AMERICAN EDUCATION: A CITIZEN'S GUIDE.* Washington, D.C.: Brookings Institution.
- Rouse, Cecilia Elena.* 1998. "Schools and Student Achievement: More Evidence from the Milwaukee Parental Choice Program." *Federal Reserve Bank of New York ECONOMIC POLICY REVIEW* 4, no. 1.
- Willis, Robert J., and Sherwin Rosen.* 1979. "Education and Self-Selection." *JOURNAL OF POLITICAL ECONOMY* 87: S7-S36.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

Standards, Information, and the Demand for Student Achievement

Richard J. Murnane and Frank Levy

Over the last eighteen years, changes in the American economy have dramatically increased the skills workers need to earn a middle-class living. However, almost half of American students now leave high school without the requisite skills. The mismatch between the growing skill demands of employers and the skills of graduating students creates a need for dramatic school improvement. Yet improvements have been slow in coming. The question is why?

In this paper, we argue that a major obstacle to higher student achievement is a lack of good information comparing achievement levels to labor market requirements—the kind of information that can come through academic standards and assessments. Without

this information, parents are unable to assess accurately the quality of their children's education.

To appreciate a parent's situation, consider the precise nature of the nation's achievement problem. When the media report that U.S. schools are in serious need of improvement, parents reasonably infer that the stories refer to U.S. schools that have collapsed. While schools in some big cities have collapsed, this is not the general pattern. The average math and reading scores of white seventeen-year-old Americans are slightly higher today than they were in the early 1970s, and the average scores of black and Hispanic seventeen-year-olds are considerably higher (see table).¹ The nation has an achievement problem not because achievement levels have fallen but because job requirements are rising much faster than achievement levels have improved.

If parents had the information to compare their children's achievement with the economy's requirements, they could see this problem and push schools for more rigorous curricula, just as they now push for anti-drug and

Richard J. Murnane is a professor at the Graduate School of Education, Harvard University; Frank Levy is the Daniel Rose Professor of Urban Economics, Department of Urban Studies and Planning, Massachusetts Institute of Technology.

TRENDS IN AVERAGE SCALE SCORES IN READING AND MATHEMATICS, BY RACE AND ETHNICITY

	Reading		Mathematics	
	1971	1994	1973	1994
Nation	285	288	304	306
White	291	296	310	312
Black	239	266	270	286
Hispanic	252	263	277	291

Source: Campbell et al. (1996).

anti-alcohol education. But without this information most parents are forced to judge schools by other standards. One such standard is the set of international test score comparisons showing that American students score lower on achievement tests than do students in many other countries. These scores appear consistent with media stories about the need for school reform. Another standard is the perception of parents—correct, in most cases—that their children are learning as much in school as they themselves did twenty-five years ago.

Taken together, the media stories and other information sources have made parents schizophrenic about the achievement problem. In the 1997 Phi Delta Kappa/Gallup poll, only 25 percent of public school parents gave the *nation's* schools a grade of A or B, yet 64 percent of parents gave the public school attended by their oldest child a grade of A or B. Parents believe that U.S. schools have problems, but the problems exist in other children's schools—a belief that has existed for more than a decade. When parents are truly this satisfied with their own children's skills, major gains in national achievement are hard to imagine. To see why, consider what school reform entails.

PRINCIPLES FOR IMPROVING SCHOOL PERFORMANCE

As we describe in Murnane and Levy (1996), organizations, including schools, that are successful in continually improving their performance recognize that the people who do the work—a group we call frontline workers—are critical resources whose skills and energies must be

engaged. Successful organizations do this by embracing five principles:

- set clear goals that all frontline workers support;
- design jobs so that frontline workers have incentives to contribute to the organization's goals and have opportunities to do so;
- provide ongoing training so that frontline workers develop the skills needed to make contributions;
- monitor progress toward goals on a regular basis; and
- persevere, even in the face of adversity, and recognize that there are no magic bullets.

Embracing these principles is difficult in any organization. It is particularly difficult in schools, because their frontline workers include not only teachers, who are on the payroll, but also students and parents, who are not. Parents and students must be drawn into any consensus on goals without resorting to the leverage that a paycheck provides. Developing consensus on the primacy of improving student achievement and on the importance of doing the hard work to achieve this goal is difficult when parents do not see these as the most urgent priorities. But how can parents understand the need for dramatic upgrading of student skills without clear evidence that their children's skills do not meet the standards needed to thrive in a changing economy?

Well-designed academic standards and assessments are not a "solution" to the achievement problem. Rather, they are a first step that makes the achievement problem concrete and visible to parents, teachers, and students. Once the problem is visible, there remains the hard, day-to-day work of making a school better.

But if standards and assessments are not sufficient for higher student achievement, they are necessary. Without the focus on achievement that they bring, other reforms—for example, charter schools, parental choice, parental involvement and professional development programs—are unlikely to have a large-scale impact. We can see both the virtues and limitations of academic standards by considering the case of the Alliance Schools Network of Texas.

THE BENEFITS AND DILEMMAS OF EXTERNAL STANDARDS: A CASE STUDY

Zavala Elementary School serves 450 children from low-income families in East Austin, Texas. Almost all of the children are Hispanic and 95 percent qualify for the free lunch program. In 1990, Zavala ranked sixty-second in student test scores out of Austin's sixty-three elementary schools. Few parents were aware of their children's low skill levels because their children received grades of A and B. Teachers gave high grades for poor work because they thought the children were not capable of better work. The situation was typical of many inner-city schools.

Texas is a state that does have mandatory tests of student achievement: the Texas Assessment of Academic Skills (TAAS). In 1991, a courageous new principal at Zavala asked a parent to stand up at a PTA meeting and explain to the assembled parents that Zavala students were scoring extremely poorly on the TAAS. When parents learned of their children's poor performance, they were outraged. Teachers were stunned; parents at Zavala had never previously questioned the quality of their children's education. The comparative test score information aroused parents, and left many of Zavala's teachers frightened.

Fortunately, Zavala had help in translating the anger generated by the test score information into productive channels. Community organizers from Texas Interfaith helped parents and teachers to build a school community committed to improving children's test scores. And the scores did improve: in 1993, 26 percent of Zavala students passed the TAAS; in 1996, 70 percent passed.

Zavala is not just the story of an outstanding principal. Parents and teachers have been able to change the culture from one of apathy to one of focused determination. When Zavala's principal was transferred to a troubled Austin middle school in 1996, the parents and teachers insisted on choosing their new principal, and selected a woman committed to continuing Zavala's strategy for meeting its goals. The school has also survived the loss of three outstanding teachers, recruited to be administrators in other Austin schools. In 1996, the percentage of Zavala

students who passed all sections of the TAAS was higher than the district average and the state average, even though the median income of Zavala families continued to be exceedingly low.

Zavala is not the only school where Texas Interfaith organizers sought to build coalitions of parents and teachers committed to improving student achievement. It is one of a growing number of schools that belong to the Alliance Schools Network: learning communities of families and school faculties committed to improving children's achievement. In the first years of the network, a critical goal was to improve the students' scores on the TAAS. For most schools in the network, this goal has been reached—a remarkable accomplishment, given the history of low educational achievement for minority group students in Texas.

The TAAS initially helped Zavala and the other Alliance Schools to embrace three of the five principles for school improvement identified earlier:

- *Set clear goals:* The information uncovered on the children's low scores provided the impetus for parent action. Improving scores on the TAAS was a well-defined goal to rally around.
- *Provide ongoing training:* For the Alliance Schools, teachers' need to raise student TAAS scores gave an urgency and focus to in-service training. This contrasts with the typical situation, in which professional development has little impact on the work teachers do with students.
- *Monitor progress:* Each year's round of test score information provided evidence of each school's success in achieving its goal.

Mastering the TAAS has been a critical step forward for the Alliance Schools. It has given the network credibility in the region and has demonstrated to participating parents and teachers that their children can learn more in school. But mastery of the TAAS has only been a first step. Members of the Alliance Schools Network are coming to understand that preparing students to succeed on the TAAS does not prepare the students to thrive in a changing economy. The TAAS can only be regarded as a minimum competency test, not a test benchmarked to the skills required to gain access to middle-class jobs.

To prepare students for success, the Alliance Schools need to set higher achievement standards and measure students' progress toward meeting these standards. However, this is a difficult task for the schools to accomplish by themselves. The efforts of the Alliance Schools would be furthered by a set of external standards benchmarked to the demands of the economy and by assessments that provide information on students' progress toward meeting these standards. This, in a nutshell, is the case for an external system of high academic standards and high-quality assessments of students' achievement.

CRITICAL SKILLS THAT SHOULD BE PART OF ACHIEVEMENT STANDARDS

To throw light on the skills that are important for students to acquire before graduating from high school, we contrast the skills used in two jobs: one paying \$7.00 per hour, the other paying close to \$20.00 per hour. Neither position requires a post-secondary-school education.²

Pickers at Sports Plus

Sports Plus is a sporting goods wholesaler that packages products made primarily in southeast Asia and distributes them to large retail stores. Pickers are the employees who package customer orders. They must know how to read and do elementary arithmetic. For instance, if Kmart orders ninety balls, and balls are packed six to each master carton, the picker must be able to figure out that fifteen master cartons are needed to fill the order. Pickers work by themselves and are expected to do just what they are asked. There are few surprises in a picker's workday. Wages start at \$6.35 per hour and extend to \$7.35.

Production Associates at Honda of America

Honda of America's Marysville, Ohio, plant manufactures Honda Accords. Production associates work in teams to assemble particular parts of the cars passing by them on assembly lines. They are expected to notice production problems and devise and implement strategies to correct them. In 1990, production associates responsible for installing heaters and blowers found that they were experiencing difficulties attaching the nuts securely to the studs

that held the blowers in place. Four associates decided to form a quality circle to diagnose and solve the problem. They wrote a brief proposal describing the problem, and management approved their working as a group on company time to solve it. The members labeled their group the Sharpshooters.

The Sharpshooters created cause-and-effect diagrams to identify possible causes, then collected data to test the various possible explanations. Using Pareto charts and histograms, they concluded that the problem stemmed from an excess accumulation of paint on the studs when the chassis passed through the paint shop.

They then pursued the source of the problem, eventually finding that it arose from the introduction of a longer stud several months ago, an engineering change made to solve another problem. Now the Sharpshooters turned to solving their original problem. They began by developing a list of possible solutions and then obtained the cooperation of the paint shop to test their solutions. Eight months after they began their work, the Sharpshooters found that their sixth proposed solution—covering the studs with masking tape before they went through the paint shop—solved the problem. The group ended its project by giving an eighteen-minute presentation to management describing how they tackled and solved the stopped blower nut problem and providing evidence supporting their solution.

Honda of America expects all of its production associates to tackle problems, just as the Sharpshooters did. The requisite skills include the ability to devise a problem-solving strategy, to develop and test hypotheses, to organize and analyze data, and to draw conclusions from the analysis. Other critical skills include the ability to communicate effectively—both orally and in writing—and to work productively in groups with people from different backgrounds. Production associates at Honda of America earn almost \$20.00 per hour in addition to an attractive health care package and other fringe benefits.

The "New Basic Skills"

Efforts by states to set standards for student achievement and to establish systems for assessing whether students

meet the standards have been plagued by controversy. A common criticism of ambitious standard-setting efforts is that states should stick to measuring the basics. But what are the basics? If the basics are the skills needed to earn \$7.00 per hour, then multiple-choice tests measuring elementary reading comprehension and the ability to divide whole numbers are sufficient. But if the basics are the skills needed to obtain and thrive in modern automobile plants and in other high-wage organizations committed to product improvement, then the list is quite different. It includes not only strong reading and math skills, but also the ability to devise and carry out problem-solving strategies, the ability to communicate effectively—both orally and in writing—and the ability to work productively in groups. These are all part of the “new basic skills” needed to thrive in today’s economy. These skills should be incorporated in the standards that all American high school graduates are expected to meet.

SOLVING THE POLITICAL AND TECHNICAL PROBLEMS

High standards for student achievement and accurate assessments of students’ progress toward meeting these standards can help schools to embrace the five principles identified earlier. If parents and teachers endorse the standards, meeting them is likely to become the chief school goal (the first principle). Assessments based on the standards can provide information on progress toward the school’s goals (the fourth principle). If it is important to teachers that students meet the standards, then the standards create incentives for teachers to focus instruction on the skills measured in the assessments (the second principle), and incentives for professional development efforts focusing on helping teachers learn to teach the critical skills (the third principle). If employers offer attractive jobs to students who meet the high standards, students have incentives to work at developing the requisite skills (the second principle). The promise of high standards is great.

Reaching agreement on academic standards, however, is difficult in a heterogeneous society. Perceptions of the skills that are important or even appropriate for

students differ. Yet the progress of states such as Kentucky, Maryland, and Vermont in setting standards shows that the challenge can be met.

The challenges of designing assessments to measure students’ progress toward high standards are also great. Aligning assessments with curriculum frameworks—the substance of what teachers are supposed to teach—is difficult. Yet close alignment is essential to getting the incentives right for teachers and students. Assessments cannot be exclusively multiple-choice tests because many critical skills—for example, writing—cannot be measured by these tests. Tests allowing open-ended responses are difficult to score reliably, as are student writing samples. Skill in one type of writing—for instance, short stories—does not accurately predict skill in another type of writing—for example, nonfictional narratives. Measuring speaking skills requires yet a different assessment methodology, as does effectiveness in working productively in groups.

While these technical problems are daunting, they are not insurmountable. The College Board has made enormous progress in developing strategies to score student writing reliably. The National Assessment of Educational Progress now incorporates many questions that require open-ended responses. In addition, technology such as video equipment offers new methods for recording and assessing student performance.

The New Standards Project, a collaboration of the Learning Research and Development Center at the University of Pittsburgh and the National Center on Education and the Economy, is a particularly promising initiative. Working with more than a dozen states and several large school districts, New Standards is building an assessment system to measure student skills in English language arts, mathematics, science, and applied learning against standards that are internationally benchmarked. The work of the New Standards group and its partners demonstrates that with sufficient resources and perseverance, great progress can be made toward developing assessments that are closely aligned with curriculum frameworks and that accurately measure students’ mastery of the skills needed to thrive in a changing economy.³

WILL TEACHING TO THE TEST BE THE ACHILLES' HEEL OF THE STANDARDS MOVEMENT?

Many teachers are opposed to standardized testing because they see conflict between the type of instruction that best educates their students and the type of instruction that produces high test scores. Teachers often use the expression "drill and kill" to describe instruction that focuses almost exclusively on preparing children to do well on particular multiple-choice tests. They argue that such instruction does little to develop useful skills.

There will always be tension between the incentives embedded in external assessments and the incentives for many teachers to do their most effective teaching. These tensions matter because external standards and assessments will contribute to improving the nation's schools only if they are palatable to effective teachers.

Evidence from Vermont and other states that are part of the New Standards Project suggests that the tensions are manageable. Teachers in these states do not like the idea of their competence being judged by their students' scores on external assessments. They point out that the students' scores depend not only on what happens in their classrooms, but also on the circumstances of children's lives outside of school. At the same time, many teachers in Vermont and other New Standards states have come to understand that preparing students to do well on the open-ended tasks included in New Standards assessments is consistent with their evolving views of good teaching. One reason New Standards assessments are gaining a following among teachers is that teachers are being involved in their design. A second reason is that the standards are relatively parsimonious; they are not a laundry list of everything a child should learn. The parsimony gives teachers considerable discretion in designing strategies to prepare students for the assessments.

NATIONAL STANDARDS OR STATE STANDARDS?

A current focus of debate is whether there should be one set of national standards and assessments or fifty state sets. There are advantages to national standards and assessments.

Most notably, they would permit parents to compare their children's skills not only with those of students in other schools in their state, but also with those of students in other states. In a country in which a great many families move from state to state, there is value in a system in which instruction throughout the country is geared toward preparing students to meet the same high standards.

There are also arguments on the other side of the ledger. Many states have made considerable progress in setting high standards and developing appropriate student assessment systems. Their efforts provide new ideas for how to measure critical student skills. Given the technical challenges of developing high-quality assessments, proponents of state standards contend that it is useful to let fifty, if not a thousand, flowers bloom.

A political argument in favor of state standards can also be made. In much of the country, states' rights and local control are highly valued, and there is considerable opposition to national standards of student achievement. Negotiations to reach agreement on a set of national standards and assessments might succeed only through a process of compromise that made the standards more like those appropriate for obtaining a job at Sports Plus than at Honda of America. This would be an enormous disservice to America's children. The evidence is not yet in on the question of whether it is possible to reach agreement on a set of national standards and assessments, but compromising on quality to achieve consensus is ill advised.

FAMILY CHOICE OR STANDARDS? A FALSE DICHOTOMY

Recent congressional debates on educational policy have evolved into a simple contest: the President's program of national tests versus the House Republicans' emphasis on school choice. This is a poor way to frame the issue.

To see why, consider the recent choice programs in Milwaukee, Cleveland, and New York City that provide low-income minority group families with opportunities to send their children to private schools. These programs demonstrate that many low-income parents want alternatives to existing urban public schools for their children. Evaluations show that many parents are more satisfied with

the private schools their children now attend under the choice programs than they were with urban public schools.

To date, evidence on the academic achievement of low-income children in choice programs is extremely limited. The most intensively studied program is the Milwaukee choice program. The math achievement scores of children who remained in the Milwaukee private schools for several years increased more—by 1 or 2 percentage points per year—than the math achievement scores of comparable students in Milwaukee public schools. There were no statistically significant differences in the rates of growth in reading achievement (Rouse forthcoming).

It is easy to understand the satisfaction of parents who see their children learning more than they did in urban public schools. Indeed, this comparison with public school student performance provides a rationale for further experimentation with choice programs for low-income

families. Yet it is important to keep in mind that by the standard of the skills needed to earn a middle-class income in a changing economy, the achievement of children in the Milwaukee choice schools is extremely low. Without dramatic improvements in achievement, children participating in the choice schools—even though they may leave school with higher achievement levels than children graduating from Milwaukee public schools—will still lack the skills to thrive in a changing economy.

Parents need to know this. A system of high standards and periodic assessments measuring whether children meet these standards would provide parents with information they need. For this reason, standards and assessments complement choice programs just as they complement the Alliance Schools Network initiative and other programs aimed at improving the academic achievement of American children.

ENDNOTES

1. The writing scores of American white and black (but not Hispanic) seventeen-year-olds were slightly lower on average in 1994 than in 1984 (the first year writing skills were measured by the National Assessment of Educational Progress) and the science scores of white (but not black or Hispanic) seventeen-year-olds were slightly lower in 1994 than in 1969.
2. Jobs at Sports Plus and Honda of America, as well as at service-sector firms, are described in detail in Murnane and Levy (1996).
3. For information on the New Standards assessments, see New Standards (1997).

REFERENCES

Campbell, Jay R., et al. 1996. REPORTS IN BRIEF: NAEP TRENDS IN ACADEMIC PROGRESS. U.S. Department of Education, National Center for Educational Statistics. October.

Murnane, Richard J., and Frank Levy. 1996. TEACHING THE NEW BASIC SKILLS. New York: Free Press.

Rouse, Cecilia Elena. Forthcoming. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." QUARTERLY JOURNAL OF ECONOMICS.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

Summary of Floor Discussion

Margaret M. McConnell

In the discussion that closed the fourth session, participants raised three concerns regarding the use of standardized tests to assess student readiness to exit the public school system. First, standardized tests might lead some teachers to “teach to the test” rather than to teach material that would address broader educational goals. Second, the potential for racial or gender bias in standardized tests could make them an unfair indicator of a student’s competence. Third, because students cannot be forced to take them, standardized tests would do little to solve the more fundamental problem of motivating students to remain in school.

The discussion opened with the suggestion that teachers in states with a long history of using standardized exit exams are more likely to teach to the test. Some participants responded that gearing instruction toward an exam is not necessarily a bad strategy. Julian Betts noted that, according to Richard Murnane, the first step toward implementing the type of standards proposed by Rudy Crew, the National Council of Teachers, and others is to make the tests themselves more interesting. To accomplish this goal, Betts suggested designing tests that assess mathematical, analytical, and writing ability through written responses rather than multiple-choice answers. Such tests would encourage students to think and synthesize information rather than to simply memorize facts. Betts commented that once tests like these were in place, educators would be free to teach to the test because the tests would reflect the school system’s underlying educational standards.

Participants then raised a second issue relating to test design: even if the format was changed so that standardized tests no longer encouraged rote memorization, the potential for gender and racial bias would remain. Betts offered some suggestions for minimizing this type of bias. He noted that test writers have already invested substantial resources in trying to write questions that tap into the general knowledge of all students. In addition, Betts suggested that the poor performance of a particular ethnic group on a standardized test does not necessarily indicate that the test is biased. The test results, he noted, are often correlated with measures of success such as earnings later in life. Thus, the scores have been measuring something important about how well our education system is working for certain groups.

Finally, participants expressed the view that the presenters’ comments on standards and testing often failed to address the more practical problems facing educators today. As one educator pointed out, the opinion that exit exams give students an incentive to work harder overlooks the fact that students can choose truancy instead; simply telling an unmotivated student that he or she has to take a test is not going to change an underlying attitude of indifference. From this educator’s point of view, we need to devise new ways of making the classroom more exciting; students must be able to see the relevance of what they are learning so that they will choose to stay in school. Unless this change occurs, the educator continued, exit exams

would do little to help save the students who are falling through the cracks. Another participant added that exit exams may be only part of the solution—requiring students to pass exams to gain post-high-school jobs may be a better way to create incentives to stay in school.

Betts responded that he was not suggesting that policymakers simply raise standards and sit back and let the problems work themselves out. Instead, he advocated using the results of standardized tests (as opposed to letter grades, which can introduce a fair amount of subjectivity to the evaluation process) as a guide in targeting more resources to the students who are experiencing the most difficulty. Betts and Derek Neal used the Chicago public

school system as an example of a program in which exit exams did seem to improve student performance. Betts and Neal also noted that while no program is going to help every student, the success of the Chicago program suggests that the implementation of exit standards is worthwhile. Asked about the use of entrance exams, Betts offered the opinion that both entrance and exit exams would be useful; however, more pressing is the underlying need to ensure that every student possesses basic mathematical and reading skills. Exit exams, in particular, seem to be an effective and direct way to assess our success in meeting educational goals.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

Summary of Roundtable Discussion

Joseph S. Tracy and Barbara L. Walter

The final session of the day, moderated by Peter Bakstansky, was a roundtable discussion addressing the issues raised in the conference papers. The discussion was led by four individuals who are actively engaged in the effort to improve schools:

- Joseph Viteritti, research professor of public administration at New York University's Wagner Graduate School of Public Policy and a coauthor with Diane Ravitch of *New Schools for a New Century: The Redesign of Urban Education*;
- Bill Andrews, executive director of National Parents Alliance, a New York City nonprofit group that sponsors "Inside Education," a cable television program designed for parents;
- Peter Flanigan, a founder of the Student/Sponsor Partnership, a nonprofit organization that provides scholarships and mentoring to inner-city youth; and
- Beth Lief, president and executive officer of New Visions for Public Schools, a nonprofit organization whose mission is to develop programs that promote better instruction, higher student achievement, and greater school accountability.

Joseph Viteritti began the session by highlighting some of the most significant findings in each of the conference papers. He went on to make general observations about the papers as a group. A fundamental lesson of the conference, he said, is that certain inner-city schools have devised a way to educate poor inner-city children successfully. The key components of these programs appear to

be more autonomy, real standards, a curriculum that meets those standards, and allocation of resources to reinforce the curriculum. Viteritti also emphasized that the school reform movement should not treat teachers and principals as factory workers who have nothing to contribute to the change process, nor should it treat parents as if they lack the ability to make intelligent choices about where their children go to school.

In closing, Viteritti commented that he sees more diversification within the U.S. education system today. He offered the following examples:

- students have greater choice in the public schools, as witnessed by the Milwaukee and Cleveland school systems;
- more private school scholarships, paid for by private philanthropists, are available to inner-city students;
- charter schools, which provide increased autonomy and accountability in using resources, are becoming more prevalent;
- parents, particularly in urban areas, are calling for more choice in schools and are rejecting a monopolistic education system; and
- principals and teachers are seeking to eliminate the regulations that seem to be running the schools.

Bill Andrews observed that the presenters touched on something that he has seen in practice: standards in and of themselves will not solve the problems in inner-city schools. Attention must also be given to the real problems

related to the governance of the public school system—most notably, the small role assigned to parent organizations relative to that of the educational administration or the unions. Andrews noted that parents are beginning to understand the magnitude of the problem—many of their children are not on the path to the “American dream,” or the American middle class. Too often, he said, they are on the path to Riker’s Island.

Andrews then commented on Catholic schooling and the reasons why Catholic schools seem to produce better educational outcomes. He argued that in Catholic schools children are not afraid in the classroom. In large, urban public schools, by contrast, many children are too frightened to learn effectively. According to Andrews, the key to educational reform is to communicate to parents that the public schools are not functioning well and to give parents the opportunity to improve the situation. Parents should also know that just pouring resources into the school system is not the answer.

Peter Flanigan spoke next, noting the huge number of troubled schools in the inner city. He cited recent articles that referred to some urban schools where only a third of students could read at their grade level as being in educational “dead zones.” Still, Flanigan observed, a number of schools in these zones do produce students that pass tests—students whose family income, parent education, risk factors, and English skills are similar to those of the failing students. So what is the ingredient present in successful schools that is absent from the others? Flanigan’s answer is competition.

Flanigan set forth two alternative models of education. In the bureaucratic, top-down, monopolistic system, the principal is tenured, has very limited freedom to deal with the curriculum, is assigned teachers who may or may not support his or her approach to running the school, and is given students who attend the school because they have no other choice. In the competitive system, by contrast, the principal has the freedom to sell his or her view of education, the flexibility to create curriculum, and the ability to choose teachers. Further, the principal and teachers

are accountable to the parents: they must say to parents, we want to teach your children, and here is how we are going to do it. Presented with the school’s offerings, the parents then make their decision.

Competition, according to Flanigan, means creating many more chartered schools and vouchers for students in poor school districts. He noted that students and parents desire bold experiments. For example, 23,000 students applied for the 1,000 scholarships that his nonprofit organization, Student/Sponsor Partnership, offered to public school students to attend private schools. In Albany, when vouchers were offered to students attending the city’s worst school, 25 percent of students accepted. The remaining students in the school did not suffer the way some thought they would. Instead, the principal was replaced, new teachers were hired, the building was painted, and the school opened in the fall with a much improved program. These examples, Flanigan suggested, show that competition is the key to reforming poor schools.

As the final lead-off speaker, Beth Lief raised three important points. First, there is no one magic bullet to solve the problems of our schools. She characterized the challenge as a search for the best partial solutions that, in combination, will improve our schools quickly. Second, standards are one part of the solution and a revolutionary part—they can change the way we think about education. According to Lief, we have moved away from the notion of a standardized Bell curve—which measures how well children perform relative to each other—to the new standards movement—which demands that all children reach a certain level of achievement. A serious problem in reaching this goal, however, is that many teachers lack the training to meet the high standards required by today’s economy, much less to prepare students to meet them.

Third, Lief emphasized that raising learning outcomes for all children will require large-scale reforms. Charter schools and more choice through vouchers are important steps, but to reach enough children we need to work at changing entire public school systems. Lief

underscored the need for better facilities, and she urged that public schools be given more power over staffing, curriculum, resources, and budgeting. At the same time, she called for greater accountability and more explicit rewards and consequences. Lief added that the consequences of failure must be felt by the adults in the system rather than by the children. Tenure needs to be examined in this context. To make sure that children are not exposed to poor teaching, the system must remove bad teachers quickly.

Following the speakers' remarks, the discussion was opened to all conference participants. The first issue raised by the group was the removal of tenure. One participant stressed the importance of preserving due process in removing tenure and noted that teacher reassignments raise complex issues. Derek Neal continued with the theme of accountability and focused on the current difficulty of rewarding good performance. He argued that unions tend to compress wage differentials so that supervisors have little latitude for rewarding individual employees for performance. Lief added that she envisioned a system designed not only to remove principals who failed, but also to nurture all principals and teachers—such a system would find and reward those leaders who enabled their students to excel.

The conversation then turned to ways to provide more choice and competition. Viteritti suggested that the fundamental rule should be that the dollar follows the child, giving poor parents the power to walk away from bad schools. The fear of job losses by teachers and others could prompt a failing school to turn itself around.

Robyn Brady offered two observations about the earlier discussion. First, she cautioned against using negative terminology to frame the debate over improving educational outcomes. For instance, calling schools that are on the failing list “dead zones” implies that there is no hope for them. Second, she noted that the topic of welfare reform had been absent from the discussion and urged participants to keep in mind that welfare changes would have a profound effect on the parents of many school children.

The session concluded with more discussion about why private schools seemed to outperform public schools. Some participants commented that rules and regulations in the public schools are more onerous than those in private schools. Another participant noted that many inner-city Catholic schools take all comers and rarely expel children, so that greater selectivity cannot explain the difference in outcomes. Finally, Andrews reiterated that inner-city public schools lack a safe environment that is conducive to learning.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.

CONFERENCE PARTICIPANTS

Gloria Ambio
Public School 38, New York

Bill Andrews
National Parents Alliance

Peter Bakstansky
Federal Reserve Bank of New York

Julian R. Betts
University of California, San Diego

Sandra E. Black
Federal Reserve Bank of New York

Katharine Bradbury
Federal Reserve Bank of Boston

Robyn D. Brady
Bankers Trust Company Foundation

Lloyd Bromberg
Federal Reserve Bank of New York

Howard Chernick
Hunter College

Matthew Cowit
*Henry Hudson Intermediate School 125,
New York*

Rudy Crew
New York City Public Schools

Doug Criscitello
New York City Independent Budget Office

Ann Davis
Marist College

Lisa Dunn
New York City Development Corporation

Randy Ann Ehrenberg
Cayuga Heights Elementary School

Ronald G. Ehrenberg
Cornell University

Raquel Fernandez
New York University

Peter N. Flanigan
Student/Sponsor Partnership

Joyce Fonollosa
Public School 45R, New York

Arlene Geiger
John Jay College

Elyse Golob
*New York City Economic Development
Corporation*

Sandra Hamburg
Committee for Economic Development

Eric A. Hanushek
University of Rochester

Caroline M. Hoxby
Harvard University

Todd Idson
Columbia University

Carol C. Ildebrando
Public School 21R, New York

David A. Jaeger
Hunter College

Pearl Kamer
Long Island Association, Inc.

Patrick Killackey
New York City Independent Budget Office

Alan B. Krueger
Princeton University

Kristin Kucsma
Drew University/Saint Peter's College

Madeline E. Lacovara
Classroom, Inc.

Seung-ho Lee
Institute of Public Administration

Frank Levy
Massachusetts Institute of Technology

Beth Lief
New Visions for Public Schools

Jenny Liu
Federal Reserve Bank of Boston

Ronnie Lowenstein
New York City Independent Budget Office

Stephen Mark
New York City Independent Budget Office

Chris Mayer
Columbia University

Jean McConnell
Rumson-Fair Haven Regional High School

Gaynor McCown
New York City Partnership, Inc.

William J. McDonough
Federal Reserve Bank of New York

Mary McInerney
Hungerford School

Reginald Melson
*Federal Reserve Bank of New York,
Buffalo Branch*

Richard J. Murnane
Harvard University

Derek Neal
University of Chicago

Maureen C. O'Brien
Intermediate School 24, New York

Dan O'Flaherty
Columbia University

Richard W. Roper
Roper Group

Cecilia Elena Rouse
Princeton University

Sonia Diaz Salcedo
Community School District 1, New York

Amy Schwartz
New York University

Harry Spence
New York City Board of Education

Leanna Stiefel
New York University

Lori L. Taylor
Federal Reserve Bank of Dallas

William Thompson
New York City Board of Education

Joseph S. Tracy
Federal Reserve Bank of New York

Joseph Viteritti
New York University

Barbara L. Walter
Federal Reserve Bank of New York

THE FEDERAL RESERVE BANK OF NEW YORK
ECONOMIC POLICY REVIEW

Single copy subscriptions to the ECONOMIC POLICY REVIEW (ISSN 0147-6580) are free. Multiple copies are available for an annual cost of \$12 for each additional subscription. Checks should be made payable in U.S. dollars to the Federal Reserve Bank of New York and sent to the Public Information Department, 33 Liberty Street, New York, NY 10045-0001 (212-720-6134). Single and multiple copies for U.S. subscribers are sent via third- and fourth-class mail.

ECONOMIC POLICY REVIEW subscribers also receive the Bank's Annual Report.

ECONOMIC POLICY REVIEW articles may be reproduced for educational or training purposes, provided that they are reprinted in full and include credit to the author, the publication, and the Bank.

Library of Congress Card Number: 77-646559