### **Conference** Proceedings



## EDUCATION IN THE 21st CENTURY:

MEETING THE CHALLENGES OF A CHANGING WORLD

Acemoglu	Kane
Barber	Katz
Betts	Kodrzycki
Bishop	Merten
Dolton	Murnane
Downes	Raymond
Figlio	Rebell
Finn	Schultz
Goldin	Simmons
Guiney	Souza
Hanushek	Wolfe
Haveman	

Editor Yolanda K. Kodrzycki

Federal Reserve Bank of Boston 47th Economic Conference June 2002

# EDUCATION IN THE 21st CENTURY:

MEETING THE CHALLENGES OF A CHANGING WORLD

Conference Proceedings

## EDUCATION IN THE 21st CENTURY:

MEETING THE CHALLENGES OF A CHANGING WORLD

<sub>Editor</sub> Yolanda K. Kodrzycki

> Federal Reserve Bank of Boston 47th Economic Conference June 2002

The production of this volume was made possible through the efforts of Ann Eggleston, Kristin Lovejoy, and Kim Underhill of the Research Department of the Federal Reserve Bank of Boston. Charts and diagrams were prepared by Heidi Furse and Fabienne Madsen, also of the Research Department. I thank these individuals for their many contributions.

Yolanda Kodrzycki Editor

### **CONTENTS**

<i>Foreword</i> Cathy E. Minehan	xi
Overview Education in the 21st Century:	
Meeting the Challenges of a Changing World	
Yolanda K. Kodrzycki	1
Address	
American Leadership in the Human Capital Century: Have the Virtues of the Past Become the Vices of the Present?	
Claudia Goldin	25
Educational Attainment as a Constraint on Economic Growth and Social Progress	
Yolanda K. Kodrzycki	37
<i>Discussion</i> Lawrence F. Katz Paulo Renato Souza	85 93
Social and Nonmarket Benefits from Education in an Advanced Economy	
Barbara L. Wolfe and Robert H. Haveman	97
Discussion	
Daron Acemoglu T. Paul Schultz	132 137

<b>Do State Governments Matter?</b> Thomas A. Downes	143
<i>Discussion</i> Julian R. Betts Michael A. Rebell	165 175
<i>Address</i> <b>The Challenge of Transformation</b> Michael Barber	181
<b>Improving Educational Quality: How to Evaluate</b> <b>Our Schools?</b> Eric A. Hanushek and Margaret E. Raymond	193
<i>Discussion</i> Peter J. Dolton Thomas J. Kane	225 237
What Is the Appropriate Role for Student Achievement Standards?	249
<i>Discussion</i> David N. Figlio Ellen Guiney	249 279 285

Panel Discussion: Policy Implications	
<b>Theories of Action for Effecting Education Reform</b> Chester E. Finn, Jr.	291
Improving Education Outcomes:	
In Colleges, Universities, and Beyond	
Alan G. Merten	295
Improving Urban Public Schools: Suggestions for Teacher Union Leaders	
Richard J. Murnane	297
An Education Support System	
Warren Simmons	304
About the Authors	311
Conference Participants	317

### Foreword

#### Cathy E. Minehan

Education is an issue that touches everyone, personally, professionally, and as citizens of our respective nations and the world. The Federal Reserve Bank of Boston has had a long involvement with education reform in Massachusetts and in Boston specifically. We do this out of a sense of community involvement, but also out of a real desire to improve the pool from which we draw a major share of our workforce. As we consider the challenges facing our country and the world, education, more so than almost anything else, is at once both at the heart of every problem and a part of every solution.

In this regard, the topic for the 47th Boston Fed conference could not have been more pressing. For some time now, we have oriented the topics for these conferences around issues central to economic growth. In recent years, we have focused on technological change, on demographic trends, on managing economic crises, and on promoting increased domestic savings. But as vital as all of these areas are, education is their equal in addressing some of the key questions of our time.

Among the topics most central to ongoing discussions at the Federal Open Market Committee are the appropriate and realistic goals for economic growth. In the late 1990s, a consensus emerged that the potential growth rate of GDP had increased, likely as a result of an increased rate of structural productivity growth. Now, with our nation coming out of a recession, there are debates about whether future productivity trends will be as strong as those during the information technology boom, or more moderate, reminiscent of what we saw in the eighties and early nineties.

The educational mix of the population is a key piece of the growthtrend puzzle. The better educated workers are, the more productive they are, and the more likely they are to provide an impetus to technological change. The more closely the mix of workforce skills matches the demands of the twenty-first century economy, the fewer the bottlenecks we will experience in pursuing our national objective of robust and sustainable growth. In terms of skill attainment, I would argue that education affects not just the quality of the workforce, but also the way in which that workforce implements new technology. Though it is difficult to prove, I believe that the interaction between capital deepening and a skilled workforce has resulted in process reengineering that has been at the heart of the jump in productivity in the United States economy in recent years. Whether productivity continues to surge, with all that this can mean for rising standards of living, depends on continued success in improving workforce skills and quality.

The Federal Reserve also has an interest in promoting education as a complement to our pursuit of sound monetary policy. We believe that a widely educated populace improves the stability of the economy. When many consumers are unable to make good use of information, or when some groups are poorly equipped to make long-range decisions concerning their financial security, free markets do not function up to their potential, and the risk of economic crises becomes greater.

We, in Boston, along with our colleague Reserve Banks, actively advance financial literacy and economic education through a number of public and community programs. One of our newest initiatives involves designing neighborhood workshops covering a range of financial literacy issues such as budgeting, homeownership, basic savings, and access to credit. To ensure the effectiveness of these programs, we need to be aware of the diverse and changing math, reasoning, language, and technological skills of the population. Only by recognizing these educational parameters can we make the appropriate choices for topics and teaching methods.

The Bank has also been active for many years in the Boston Private Industry Council, working to improve both the employment prospects of the graduates of our local schools and our own workforce. One of our recent initiatives, "Classroom at the Workplace," offers high school students an opportunity to improve their literacy and math skills as part of their summer jobs. These summer employees attend daily 90-minute classes designed by teachers from the Boston public schools that take place on our site. Begun as a small effort by our Bank, Verizon, and Gillette three summers ago, this program now involves 20 employers and 300 students from Boston public schools. These numbers mean that we stand a good chance of helping about a third of the Boston seniors who have not yet passed the high-stakes test that is a state graduation requirement in 2003.

Many of the decisions that we, as individuals, make about education seem very localized, not just the personal decisions we make about schooling for our families, but also the stances we take as voters and public citizens. For those of us living in Massachusetts, these decisions include whether we support or reject tax-limitation measures that affect school funding, and whether we choose to support or reject standardized testing as a high school graduation requirement. Across the United States, state by state, similar deliberations are taking place that alter public education. But these decisions are also being replicated throughout the world in developed and developing countries alike.

Recently, James Wolfensohn of the World Bank spoke of a new initiative to implement reforms in education (and two other key areas aimed at jump-starting development) in 15 or 20 developing countries. His concise assessment of the related challenges—time and scale—struck me as at the heart of all of the efforts we need to be making locally in education reform. No matter where you are in the world, education is a vital component of development, and one that defies easy solutions as we seek to leverage its contributions to economic growth.

Thus, I believe it is a hopeful sign that the current wave of worldwide education reforms involves rigorous economic analysis to an unprecedented extent. It goes without saying that this is a welcome development to an institution such as the Federal Reserve that relies on and values economic research. But it also follows the tradition of microeconomists making noteworthy contributions to public policy in a variety of important sectors of the economy. Four decades ago, Kenneth Arrow, economist and, ultimately, Nobel laureate, published an analysis of healthcare markets in the American Economic Review. At the time, it was radical to view healthcare as operating within standard models of economic behavior and pathbreaking to analyze how imperfect consumer information affected healthcare delivery. Today, Arrow's insights remain at the core of how policymakers view healthcare, just as Ronald Coase's scholarship guides environmental and natural resource policies, and as Alfred Kahn's initiatives have transformed transportation markets. Will the application of economic modeling provide us with insights that are useful in creating better education systems-systems that meet the challenges of time and scale? I believe that the answer is ves, and that this conference, bringing together so many different backgrounds and fields of expertise, helps provide some new insights.

December 2002

Jacky E. Hinchon

Cathy E. Minehan President and Chief Executive Officer Federal Reserve Bank of Boston

### Overview

## EDUCATION IN THE 21ST CENTURY: MEETING THE CHALLENGES OF A CHANGING WORLD

### Yolanda K. Kodrzycki\*

During the twentieth century, the United States was a world leader in raising the educational attainment of its population. This important achievement contributed to national productivity growth and extended economic opportunity to formerly disadvantaged groups in society. Now, at the beginning of the twenty-first century, U.S. institutions of higher learning retain an excellent reputation for quality. Less confidence exists, however, in the educational system's ability to meet broad economic and social objectives adequately. This uncertainty stems in part from the shifting global economy and the evolving nature of employment. These doubts also reflect the legacy of widening income inequality over the past quarter-century. These concerns have sparked both federal and state legislation to reform elementary and secondary schooling.

The Boston Fed's 47th annual conference brought together experts from a variety of perspectives to analyze current institutional and financial arrangements in the area of education, with the goal of identifying the nature of the shortcomings and appropriate ameliorative actions. Although the primary focus was on the U.S. educational system, the Bank welcomed international perspectives. The experience of other nations provided evidence on the degree to which educational challenges are being driven by changes in the worldwide economy, and offered insights on the strengths and weaknesses of alternative educational systems.

<sup>\*</sup>Assistant Vice President and Economist, Federal Reserve Bank of Boston.

#### **CONFERENCE THEMES**

A central theme of the conference was that the U.S. educational system is in the process of being restructured. The key debate is no longer about funding for education. It is about how to change institutions and incentives so as to bring about better educational outcomes.

Dissatisfaction with the current education system in the United States was ubiquitous among conference participants. To varying degrees, all claimed that the performance of the average student should be improved, that the educational attainment of low-income and minority students must be raised from current unacceptable levels, or that greater attention should be placed on developing high-end talent. As a result of their concerns, participants generally welcomed the greater emphasis that public and private officials are placing on improving schools.

Conference participants agreed that education is increasingly important in determining individuals' earnings potential. They also agreed that the total benefits to society from education are greater than the sum of what individuals earn as a result of their educational attainment. Participants reached a consensus that these links between personal and social well-being and education need to be better communicated to the U.S. populace.

Relative to foreign populations, the U.S. population, on average, is highly educated in terms of years of schooling. However, the average U.S. high school or middle school student does not score highly on international standardized tests. As a response, some participants would concentrate on increasing academic achievement for a given number of years of schooling. Others would focus more on increasing the fraction of the population that completes secondary and higher education, especially since gains in educational attainment have slowed among younger cohorts.

Recent education-related reforms in the United States have had two key thrusts. "Standards-based reforms" involve establishing performance benchmarks for students and schools and holding them accountable for their performance. "Choice" involves providing expanded alternatives to traditional public schooling, such as through vouchers and charter schools. In addition, over the last several decades, states have implemented a variety of changes in school financing in response to voter and legislative actions and judicial decrees. These reforms to school funding formulas are ongoing.

In the case of standards-based reforms, two papers presented at the conference point to evidence of likely improvement in academic achievement. Nevertheless, for a variety of technical and philosophical reasons, attendees differed in their assessments of standards-based reforms. Evidence on the efficacy of vouchers and charter schools is still quite limited, given their small scale and relative newness. Finally, on the whole, research indicates that the changes in school funding implemented by various U.S. states have resulted in only limited changes in student performance.

### STANDARDS-BASED REFORMS: SMALL STEPS IN THE RIGHT DIRECTION?

Conference participants warned policymakers and the public not to declare victory in meeting the challenge of educational reform. Many expressed the view that standards-based policy changes to date represent comparatively small steps, albeit in the right direction. Others warned of possible negative implications from the standards movement.

Those who supported the general thrust of standards-based reform pointed to its potential to raise academic achievement. Nevertheless, some adherents of performance benchmarks also cited its inadequacies. Remedies for these problems include raising standards further, refining how test scores are used, or making additional, complementary investments in educational reform.

Participants indicated that, in some states, the new standardized tests either are not rigorous enough to have an effect on student performance or are not sufficiently oriented toward the skills needed to succeed in twenty-first-century labor markets. Furthermore, in most cases, states are not using the information from tests in ways that provide accurate assessments of schools. Teachers often are not receiving information on student test scores in a timely fashion and, in any case, may not have the training or resources to improve their teaching. Participants also advocated for additional institutions to join in the educational reform movement. These institutions include teachers' unions, colleges and universities, and social-services providers. Finally, one of the panelists argued that standards-based reforms were more effective when combined with greater choice, which has been only a minor feature of the changes implemented in the United States. All in all, many attendees agreed that countries on the forefront of standards-based reforms, such as the United States and the United Kingdom, face the ongoing challenge of transforming their educational systems.

Those who appeared more skeptical of the current wave of standards-based reforms were inclined to bring up the tradeoffs associated with any set of incentives. They noted that education encompasses multiple goals, some of which are not reflected in standardized tests. For example, some of the strengths of the U.S. economy—such as an entrepreneurial workforce—can perhaps be traced to aspects of its educational system. Speakers warned that testing efforts run the risk of diverting resources from some sets of students to others, in ways that may not be transparent or desirable. They also pointed to some conflicts in the incentives created by state educational reforms versus those included in the federal No Child Left Behind legislation. How these differences are resolved will have a bearing on the success of reforms in the United States.

Underlying some of the differences of opinion were very different philosophies on the merits of having government-imposed standards for education. One prominent educational reformer noted that educational systems traditionally have been based on implicit standards. He argued that explicit standards are superior because they are more transparent. Educators know what is expected of them, and they can design instructional systems that move toward these goals. On the other hand, other participants expressed the view that explicit standards were inherently harmful. For example, one member of the audience likened the situation to central planning in the Soviet Union. When steel producers were judged on tonnage, they reduced quality as they increased the quantity of production. In this speaker's view, for a variety of markets (including both steel and education), "the only way to make progress is by relying on competitive mechanisms where the customers take their business to the firms with the better products." In the context of the current standards-based reforms, another observer saw perverse repercussions from calls for further research on educational effectiveness. Such investigations could lead to testing "beyond the realm of good policy." He commented, "You don't fatten a pig by weighing it."

### THE NEED FOR GREATER SUPPORT FOR URBAN SCHOOLCHILDREN

Conference participants agreed that recent efforts to narrow the educational attainment gap between children from wealthy and poor communities have, on the whole, met with limited success. And although the recent interest in standards, accountability, and expanded school choice has been motivated by the view that increased school funding has only limited effects, the presenters advocated a range of policies that arguably would require higher levels of funding for schools in communities with high concentrations of poor and immigrant families. Such schools increasingly are found in large cities.

At a minimum, the solution to educational disparities was said to involve shifting a greater share of overall education funding to elementary and secondary schools in poor areas. However, most of the discussion implicitly seemed to support increased funding for such schools without offsetting reductions in funding for schools in wealthier areas. Participants emphasized that urban schoolchildren face a multitude of problems outside of the schools. They advocated policies that would supplement the services provided during the regular school day (or regular school year) or that would expose urban children to environments outside their inner-city neighborhoods. Moreover, to the extent that the low college-attendance rates among students from poor and minority families reflect barriers to financing higher education, the solutions were said to lie either in greater public subsidies for higher education or in greater resources for financial aid.

## Limitations in the Evidence on What Works and What $\ensuremath{\text{Doesn't}}$

Conference participants emphasized that education researchers often cannot provide unequivocal answers to what may appear as basic questions to policymakers. Standards and school choice are relatively recent innovations. Some of the effects may not be apparent until these efforts achieve a certain scale.

Beyond the inherent difficulty of analyzing new educational structures, conference participants agreed that education research does not yet have definitive answers to underlying questions governing resource allocation such as "How do we produce better-educated individuals?" and "How large are the societal benefits of better education?" For example, conference attendees had an animated discussion of what is more important for disadvantaged children: increasing the quantity or raising the quality of the education they receive. Even those who support "quantity" over "quality" may favor different mixes of emphasis among preschool education, summer programs, and access to higher education.

Researchers struggle with even the most basic questions because the process of producing better-educated individuals is complicated, involving student effort, schools, and family and neighborhood influences. Similarly, the production of societal benefits such as improvements in health or reductions in crime also involves a complex mix of inputs, of which schools are only one component.

Participants cautioned that research on education must be presented and interpreted in a way that reflects the preliminary state of many of the findings. This attitude was reflected in both the formal and the informal exchanges. Speakers often disagreed about the magnitudes of the effects of various policies, but they tended to support a blend of approaches to educational reform rather than promoting the exclusive use of a single approach. As testimony to the complexity of educational issues, one veteran attendee of Boston Fed conferences remarked at the conclusion of the conference that this year's speakers seemed more humble and open to discussion than is often the case.

### WHAT PRODUCED THE HUMAN CAPITAL CENTURY?

**Claudia Goldin** opened the conference by reflecting on educational structures in the "human capital century." The idea that the wealth of a nation is embodied in its people was first voiced in the United States at the beginning of the twentieth century. By the end of the century, the recognition that education is essential for technology adoption and economic growth was universal. Over 100 nations of the world currently provide secondary school enrollment data, and almost all of these countries have higher enrollment rates than the United States did in 1900.

The United States made rapid strides in secondary and higher education in the first half of the twentieth century—despite the arrival of many poor immigrants from other parts of the world. By the mid-1950s, almost 80 percent of 15- to 19-year-olds in the United States were enrolled in school. In contrast, most European nations had general school enrollment rates of less than 30 percent for this age group. Even including the relatively high technical school attendance in Europe, a wide gap existed compared to enrollment rates in the United States.

Goldin argues in her address that the early support for mass secondary education and expanded higher education in the United States was consistent with the economic opportunities of the technologically dynamic, socially open, and geographically mobile New World setting. She identifies various "virtues" of secondary education in the United States that promoted mass education. For example, U.S. secondary schools have been publicly funded and managed by small, fiscally independent districts. Goldin argues that small districts are a virtue because taxpayers who are relatively homogeneous with respect to characteristics such as income, ethnicity, religion, and cultural values are more likely to support education than taxpayers from larger districts (or, as in the case of Europe, nations), where preferences for public goods tend to be more disparate.

Goldin further characterizes twentieth-century U.S. secondary schools as "open and forgiving," secular in control, and gender-neutral. Students could enroll regardless of age, social status, previous school record, religion, or sex, which encouraged school attendance among populations who might have been excluded in a more rigid structure.

Finally, relative to the situation in Europe, the curriculum in U.S. secondary schools was "academic yet practical." Students were exposed to a broad base of knowledge that could be applied in a wide variety of occupations. In Europe, by contrast, all but an elite group of youths were channeled into an industrial or specific vocational track that precluded their access to higher education or high-end professions.

Now, at the dawn of the twenty-first century, some of these American education "virtues" are viewed as possible "vices." Popular support for publicly funded alternatives to traditional public education has grown. Indeed, the Supreme Court ruling in *Zelman* v. *Simmons Harris*, handed down just one week after the conference, provides further impetus for allowing families to use publicly funded vouchers in private schools. Small, fiscally independent school districts, once seen as a structure that promoted greater spending on education, now are being viewed as a source of serious funding inequities. Educational standards and sanctions for students and schools that do not pass are viewed as potential remedies for the lack of accountability brought on by the open and forgiving systems of the past. Thus, Goldin concludes that an entirely new set of "virtues" could emerge in the twenty-first century.

### The Relationship between Economic and Social Progress and Education

Yolanda Kodrzycki addresses the links between education and the economy in the conference's first presentation. Kodrzycki concludes that improving the quality of U.S. education should be of rising concern. In addition, as racial and ethnic minority groups account for a growing share of the U.S. population, improving their educational opportunities goes hand-in-hand with overall economic growth objectives.

Examining the U.S. evidence, Kodrzycki shows that overall high school and college completion rates have risen considerably since 1970, but that progress among younger cohorts has slowed. Although the United States has the highest international ranking for average number of years of schooling completed, average scores on standardized tests administered to secondary school students are not in the top half of the international distribution and have not improved in recent years. Kodrzycki interprets these test score findings as evidence of mediocre quality of schooling for the typical U.S. student and predicts that the lack of improvement in education could constrain productivity growth in coming decades.

The educational attainment of minority groups is of increasing importance because their share of the population is rising. The most dramatic population increase has been among Hispanics, who now constitute over 15 percent of young adults, compared to only 5 percent three decades ago. Among young adults, the gap between black and white high school completion rates has been closed, but a large gap continues to exist in college completion rates. School completion rates among Hispanics lag far behind for both high school and college, owing in part to large numbers of recent immigrants. Furthermore, at comparable levels of education, black and Hispanic minorities perform worse by various measures and have fewer classroom resources than whites.

Kodrzycki performs simulation exercises to determine how much of the gap in earnings between whites and minority groups is due to educational differences. She finds that, for full-time male earners, onefifth to one-third of the earnings gap is due to minority groups' having fewer years of schooling than whites. For females, the deficit in the amount of schooling accounts for roughly one-half of the earnings gap. The remainder of the observed wage gaps for full-time earners is attributable to minorities' earning less than whites for comparable years of education. On the basis of the evidence concerning indicators such as test scores, computer and Internet access, and literacy, Kodrzycki concludes that the lower earnings reflect the fact that blacks and Hispanics receive lower-quality education than whites—or, at least, that the education they receive does not make up for any deficiencies arising from family resources and neighborhood influences.

Finally, Kodrzycki examines the evidence on shortfalls of talent in scientific and technical fields. These concerns emerge periodically because the demand for workers trained in engineering, information technology, and similar occupations tends to spike upward abruptly in response to changes in technology or government policies. The supply of such workers inevitably responds with a lag, given the length of time required for education and training, causing a temporary shortage. Looking ahead to the coming decades, projections for only modest growth in the number of college graduates in the United States imply some constraints in filling positions, even if students respond to market signals when choosing their college major. Thus, Kodrzycki concludes that mechanisms to retrain the adult workforce appear to deserve greater attention than in the past.

In his discussion of Kodrzycki's paper, **Lawrence Katz** notes that the slow growth in the supply of college-equivalent workers in the United States during the last two decades stands in sharp contrast to the increases earlier in the twentieth century and has had a major impact on wage inequality. Other countries with decelerations in the rate of educational advance in recent decades—such as the United Kingdom and Canada—also have experienced substantial increases in educational wage differentials. By contrast, countries with continued rapid expansion in educational attainment—France, the Netherlands, and Germany—have not.

Students from lower-income and minority families account for much of the slowdown in U.S. college enrollment and completion rates, and these types of families increasingly are located in inner cities. Katz suggests that programs to assist low-income and minority families in moving to other locations with higher school quality, greater safety from crime, and supervised after-school activities should be considered important complements to educational policies that are designed to improve human capital development.

Finally, Katz addresses Kodrzycki's simulation results showing that the preponderance of black–white wage differentials occurs within education groups. He cautions that this finding does not imply lower returns to education for minorities than for whites. Instead, it largely reflects developmental deficits associated with differential family, neighborhood, and school resources, as well as lingering racial stigma and labor market discrimination. Katz views policies to raise the quantity of schooling received by minorities as the single most important lever for improving their economic status.

The second discussant, **Paulo Renato Souza**, addresses education in economic development, focusing on the example of Brazil, where he was serving as Minister of Education at the time of the conference. From 1900 until 1975, Brazil's rate of economic growth was second only to Japan's. Yet, despite enjoying the reputation of having the best higher education system in Latin America, Brazil had very high illiteracy and dropout rates, especially in the poorest sections of the country and among blacks.

With the growing importance of knowledge as the basis of economic growth in the latter part of the twentieth century, Brazil determined that it could no longer base its economic policies on abundant natural resources and cheap, uneducated labor. The nation now recognizes that if it is to maximize economic growth, its education system must promote the ability to learn and provide its citizens with opportunities for lifelong learning.

To further this goal, during the 1990s, the national government of Brazil implemented the Bolsa-Escola debit card program, which provides grants to low-income families whose children are enrolled in school. The Brazilian government also revamped curriculums and the system for evaluating schools. Between 1994 and 2000, the overall elementary school enrollment rate increased from 87 percent to 96 percent, and the differentials by income and racial group narrowed significantly. The percentage of students repeating grades fell, allowing greater percentages to pursue secondary education before entering the workforce. In the discussion period, it was noted that Mexico has successfully implemented a program similar to the Brazilian Bolsa-Escola card to boost attendance among poor students in rural areas.

### BEYOND LABOR MARKET EARNINGS: THE SOCIAL RETURNS TO EDUCATION

Determining the appropriate level of investment in education requires going beyond the earnings effects that have been the traditional focus of economics literature. Accordingly, **Barbara Wolfe** and **Robert Haveman** catalog and estimate the social returns to education. For example, greater parental education is associated with greater education for children and improved health of children. People with more education tend to make more efficient consumer choices, devote more resources to charitable giving, and commit less crime.

Economists have used a variety of techniques to isolate the effects of

schooling on labor market earnings, independent of additional compounding factors such as a student's ability, drive, or family influences. The consensus from this research is that an added year of education yields a rate of return between 7 percent and 9 percent. Adding in the full range of benefits from education, Wolfe and Haveman estimate that the total social returns may be double the conventional estimates. Their conclusion implies that investments in education should be increased from current levels.

The added social returns analyzed by Wolfe and Haveman fall into either or both of two categories. Private nonmarket returns are the nonmonetary benefits that families receive from education, and externalities are the benefits received by others in society. These types of benefits are not directly valued in the marketplace. To arrive at estimates, Wolfe and Haveman appeal to the economic theory that people combine efficient mixes of "inputs" (such as financial resources, education, and so forth) in achieving desired "outcomes" (such as, for example, improved health status for oneself or one's children, or increased education for one's children).

The authors then estimate the implicit marginal value of schooling by drawing on empirical studies measuring how "productive" a dollar of financial resources and a year of education are in achieving different social outcomes. On the basis of the existing literature, which encompasses only some of the nonmarket returns and externalities cataloged in their paper, Wolfe and Haveman conclude that the total social returns to schooling may be as great as 14 percent to 18 percent.

Wolfe and Haveman observe that developed countries tend to devote about 5 percent to 7 percent of their GDP to education, including both private and government spending as well as the forgone earnings of college and university students. Since few other investments seem able to claim returns as large as their estimates for education, the share of societal resources devoted to education should likely be increased.

Wolfe and Haveman caution, however, that their research does not indicate how the extra spending should be allocated between the private and public sectors. To reach such a conclusion requires determining what share of the benefits individuals receive, versus how much constitutes an externality received by society at large. It also requires forming a judgment on the need for government intervention to alleviate the constraints lower-income families face in paying for education.

The discussants agree that investments in education should be guided by comprehensive measures of the returns and that the total social returns to education exceed the labor market returns that have been the traditional focus of economic studies. Each focuses his remarks on ways to improve the measurement of social returns and endorses the need for a next generation of research along these lines. As an introduction to his discussion, **Daron Acemoglu** raises two key questions related to determining the appropriate level of investment in education. First, based on the external returns to education, should governments intervene more in this sector than we observe today? Second, has the overall societal return to education increased over time, mirroring the trend in private pecuniary returns? Acemoglu concludes that these questions have not yet been answered adequately.

Acemoglu points out that the studies employed by Wolfe and Haveman are based on the ordinary least squares (OLS) methodology and, therefore, do not establish the true causal link between education and outcomes. Individuals who obtain relatively high levels of education differ in their family and social background from individuals who receive less education, and these background variables likely contribute to their observed choices of what to consume, where to live, how to raise their families, and so forth. It is misleading, he says, to attribute all of the observed differences to their higher levels of education.

Acemoglu illustrates the importance of methodology by using an example from his own research of the spillovers from education in local labor markets. Studies using the OLS methodology conclude that the average worker is more productive, and is therefore paid more, in locations with a high concentration of highly educated workers. The implication is that the presence of these educated workers has beneficial effects on the working population at large, such as through more pervasive adoption of technologies and organizational arrangements that enhance productivity. However, when Acemoglu applies instrumental variable techniques to the same question, he finds that the spillover effects in local labor markets from highly educated workers to other workers are minimal at best.

Acemoglu acknowledges that his findings on local labor markets do not preclude the importance of other social benefits from education. For example, recent research using appropriate instrumental variable methodology finds that individuals who obtain more education, because compulsory schooling laws preclude them from dropping out of school, are less likely to commit crime.

**Paul Schultz** also focuses on methodological issues in his discussion. A technical assumption underlying Wolfe and Haveman's approach is that more highly educated individuals are more efficient in producing outcomes, but that they do not differ from less-educated individuals in the "production techniques" they employ. Schultz cites studies on agriculture showing that more-educated farmers are more productive, in part, because they use production techniques different from those used by less-educated farmers.

Schultz argues that similar mechanisms may be at work in the child-rearing context. More-educated mothers may manage to produce healthier children by substituting other inputs for their time. In computing the social benefits of mothers' education, one must subtract the cost of these added inputs.

More generally, Schultz calls for deeper research into the technology of production of nonmarket goods: How do educated parents allocate their time? What activities benefit and suffer as a result?

### **DOES FUNDING MATTER?**

Thomas Downes reviews the evidence on how state and local financing reforms have affected educational quality. His study encompasses three sets of reforms: court-mandated changes in the allocation of state aid to local school districts; voter-imposed limitations on local taxes used for education; and state funding of alternative educational institutions, such as charter schools. On the whole, these reforms have served to increase the states' share of elementary and secondary school funding and may, therefore, have provided more equal funding across school districts within states.

Downes concludes that finance reforms implemented in response to court orders seem to have little, if any, impact on the distribution of student performance. Tax and expenditure limits appear to be associated with some decline in average mathematics scores and (at least in one study) an increase in dropout rates, but he finds no discernible changes in the distribution of student performance across school districts.

Compared to the evidence on other finance reforms, the literature on charter schools remains quite limited because they account for only about 1 percent of total student enrollment nationwide. Students attending charter schools have been found to experience an initial, temporary decline in test scores, which is consistent with the general findings on students who change schools. The presence of charter schools does not appear to change the performance of students in traditional schools, although some evidence suggests that charter schools have a positive effect when they provide a threshold level of competition. Because the effects of charter schools seem to be different, depending on how long they have been in operation and how large a share of the local school market they account for, Downes concludes that the long-run effects of competition have yet to be evaluated.

In reviewing the school finance literature, Downes distinguishes between two strands of research: studies of reforms in particular states and national comparisons of generic (or "canonical") reforms. Downes observes that each type of research has idiosyncratic strengths and weaknesses. For example, policymakers interested in the effects of reforms in California would benefit from studying the consequences of the restrictions on local funding of education that have grown out of the 1976 *Serrano* v. *Priest* court case and the results of the limitations on property taxation and per pupil spending imposed by the 1978 passage of Proposition 13. Because of unique circumstances relating to both the reforms and the economy in California, however, the conclusions would not necessarily provide useful indicators for other states. On the other hand, comparisons of student test performance in states that had and had not undergone finance reforms could provide estimates of the average impacts of such categories of reforms but should not be construed as evaluating exactly the same policy changes in each state. Downes concludes that state-level and national-level analyses should be used in concert to guide policymakers.

Both discussants express the view that, to date, the research on school funding has fallen short of what policymakers need. **Julian Betts** prefaces his remarks by emphasizing the difficulty of drawing definitive conclusions about what causes educational outcomes. While disparities in school finances may matter, so do disparities in home and neighborhood environments, as well as hard-to-capture differences among school districts—such as the quality of local school administrators and the attitudes of local residents. In his view, these other factors remain influential, whatever equalization may occur as a result of fiscal reforms.

Betts argues that even careful econometric studies may mistakenly attribute changes in student achievement to changes in the financial support for schools, while ignoring simultaneous developments that differ across states or school districts. For example, concern about poor student achievement may lead both to lawsuits that result in changes in financial resources and to increased parental involvement in the schools. A similar problem exists in trying to assess the impact of charter schools: Concern about student achievement that leads to the establishment of more charter schools may also lead to the hiring of reform-minded administrators who take steps to improve all public schools. Whatever change occurs in public-school student achievement should not be attributed only to competition from charter schools. Because state-level analyses inevitably fail to capture some relevant details, Betts recommends further pursuit of district-level studies.

**Michael Rebell** provides a context for his remarks by observing that suburbanization in post–World War II America brought increased economic segregation, leading to unprecedented disparities in financial resources across school districts. As dramatized three decades ago in the U.S. Supreme Court case of *Rodriguez* v. *San Antonio Independent School District*, poor school districts could fall far short of matching the school funding provided by nearby wealthy school districts, even if they were willing to levy relatively high property tax rates. Since the *Rodriguez* decision held that the federal courts could not provide a remedy, school-funding cases have fallen to individual state courts.

Rebell interprets the economics literature reviewed by Downes as saying that "money doesn't matter," since, on average, states that were subject to court decrees on school financing did not show any convergence in the academic performance of students from rich and poor districts. He finds this conclusion unhelpful, if not misleading. In many states, court orders had limited implications. In some cases, they were ignored by state legislatures, while in other cases, they pertained only to certain forms of education spending. The national studies performed to date fail to distinguish between such circumstances. A more useful approach, according to Rebell, is to perform comparative case studies of individual states. This methodology would likely uncover legal strategies that are effective in bringing about equalization of educational resources and performance.

On the whole, the session on educational funding appeared to result in participants' reaching two different—though not necessarily conflicting—conclusions. Those inclined to believe that "money must matter" called for further study of how to allocate school budgets more efficiently and for broad dissemination of the findings. For example, what is known about the efficacy of lengthening the school year or of alternative ways of investing in professional development for teachers? Others more inclined to believe "money doesn't matter much" voiced support for experimenting with standards-based reforms or school choice.

#### LONGER-TERM GOALS FOR EDUCATION REFORM

In his address, **Michael Barber** provides perspectives on education reforms in the United Kingdom, which he has overseen on behalf of the Blair government. Since the late 1980s, the United Kingdom has put in place a framework for continuous improvement of education. Many of the measures are similar to reforms in the United States. The U.K. framework includes setting high standards through a national curriculum and school inspections, substantial budget allocation authority for individual schools, readily available data to enable schools to compare their own performance against those of other schools, and expanded investments in instructors' professional development.

Signs of success to date include rising scores on international standardized mathematics and literacy tests, such as the Program for International Student Assessment (PISA). However, the Blair government remains cognizant of a long list of remaining challenges. These include making better progress at the secondary school level (such as lowering dropout rates and creating more effective vocational education programs and links to out-of-school learning opportunities), offering higher-quality programs for the most-talented students, increasing access to university education for students from lower socioeconomic groups, and developing the leadership talent of head teachers.

Barber draws an analogy between the ongoing efforts at education reform and the mission of explorers Lewis and Clark in the early 1800s to discover a route from the East Coast to the West Coast of the United States. In Barber's words: "I feel as though we've reached Kentucky, but we don't know what's beyond the Mississippi." The remainder of Barber's remarks address his longer-term vision for transforming the educational system of the United Kingdom.

In Barber's mind, the transformation of the education system must be directed at achieving two goals simultaneously: having the most talented workforce possible and improving the equity of educational outcomes. Essential in this process is moving to a system of informed professional judgment, whereby teachers have access to high-quality data on student performance and teaching practices, and where their teaching is driven by what these data tell them. Under such a system, the process of teaching would be re-engineered, with time reallocated toward activities such as professional growth, planning, and mentoring. Schools might choose to be combined into flexible networks that share innovations and services with one another. Educational outcomes would be transparent to taxpayers, to students, and to their families. Moreover, as schools become genuinely responsive to the learning needs and aspirations of individual pupils and their families, Barber envisions less need for the kinds of formal accountability systems that are currently being developed.

#### How to Assess School Performance?

**Eric Hanushek** and **Margaret Raymond** evaluate the U.S. experience in setting up accountability systems for schools and school districts. As the authors point out, as of 1996, only 10 states had active accountability systems, while by 2000, just 13 states had yet to introduce such systems. Under the federal No Child Left Behind Act of 2001, all states must move in this direction. Hanushek and Raymond review the diversity of accountability systems across states and explore their incentive effects.

The authors express concern that state performance benchmarks often emphasize process and input measures that are relatively easy to change but that have been found to bear little relationship to student achievement. In the authors' words: "We know how to order more computers or to deliver new programs; they are the low-hanging fruit on the accountability tree."

Even when states use performance benchmarks such as standardized test scores, which Hanushek and Raymond claim are closely linked to student achievement, states tend to report the results in ways that prevent an accurate assessment of how well or poorly schools are performing. States most commonly issue what Hanushek and Raymond refer to as "status-change measures." For example, such measures indicate the change in the average test score for a particular grade in a particular school or school district from one year to another. The problem with this approach is that the students in, say, third grade in one year are different from the students in third grade the next year. Improvement in average scores may simply reflect a better draw of students (from more advantaged backgrounds, for example), rather than any overall improvement in schooling. A superior approach, they argue, would involve tracking individual students over time and aggregating these year-by-year changes into an overall summary for the school or school district. Only four states currently adopt this approach, which is much more demanding from the perspective of data requirements.

Some previous studies referenced by Hanushek and Raymond examine changes in student performance and other outcomes after individual states introduced accountability systems. Hanushek and Raymond present the first-ever attempt at measuring whether states that introduce accountability systems show more marked improvement in student performance than those that do not. Using National Assessment of Educational Progress (NAEP) mathematics scores for two student cohorts in the 1990s, the authors find that the presence of some form of accountability is associated with an increase in state NAEP scores. They also find only weak support for the view that states that merely issue "report cards" on schools see less increase in student test scores than states with a system that has some form of reward (sanction) for good (poor) performance.

Some critics of state-standardized testing argue that it provides incentives to place greater numbers of students into special education programs so as to exclude them from the tests and thereby boost reported average scores. To the contrary, Hanushek and Raymond provide statistical evidence that although states introducing standardized testing did increase special education placements, these increases were not out of line with the nationwide trend during the 1990s.

In his commentary, **Peter Dolton** argues that designing incentives to achieve education goals is inherently difficult. For one thing, education encompasses multiple goals—not just achieving higher test scores. If educators are expected to devote effort to important but hard-to-measure goals—such as fostering the emotional growth of children and preparing them for their eventual social responsibilities—then the incentives to achieve measurable goals must be weakened. Another issue is that teachers are responsible to multiple stakeholders, including school heads, education authorities, parents, taxpayers, and others. To the extent these groups have competing objectives, the incentives teachers face with the introduction of an accountability system are inevitably weakened.

Dolton argues that the conditions needed for accountability systems to provide an efficient allocation of educational resources do not square with reality. In particular, effective accountability requires that all the consumers of education have the power to influence educational priorities as well as the means to choose alternative providers of education in a competitive environment. What happens, Dolton asks, if the voices of more affluent and more highly educated parents prevail over those of other "consumers"? What happens if these influential parents choose to exit the public school system in favor of private schools, rather than continuing to voice their concerns?

Although the introduction of greater accountability has been associated with improvements in standardized test scores in both the United States and the United Kingdom, Dolton argues that little if anything is known about its other possible effects. For example, does accountability result in greater expenditure of public or family resources? Does it result in school resources being reallocated away from top- and bottomperforming students, and more toward those students who are at the threshold of passing the tests? Does accountability ultimately result in less progress in meeting long-term objectives, such as a better citizenry or a labor force with more transferable skills?

Thomas Kane emphasizes that school test scores provide imprecise signals about how well schools are performing in a given year. Scores can be affected by transient events, such as poor classroom chemistry in a given year or a school-wide disruption on the day of the test. As a consequence, average school test scores exhibit relatively weak correlation from year to year, especially in the case of small schools. The problem of imprecision becomes even worse when states base their evaluations on changes in performance over time. Furthermore, the variation in scores across schools is much smaller than variation across students within schools, casting doubt on the advisability of interpreting test scores as measures of how well or poorly different schools are performing.

Given the inherent imprecision in measuring performance, some commentators have questioned whether state accountability systems might err in rewarding high-ranking schools too much. Kane provides evidence to the contrary. In California, which has a relatively generous award system for schools and faculty that achieve improvement in test-score performance, the awards are at most only one-tenth of the payoff the students can expect to receive in the labor market as a result of greater learning. Thus, in a sense, the inexactness of test scores is already taken into account in establishing only small incentives for educators.

On the other hand, Kane expresses concerns about possible inconsistencies between existing state accountability standards and those being introduced by the No Child Left Behind Act. States tend to sanction or reward schools based on changes in performance over time, but under the federal legislation, schools will face sanctions if any racial or ethnic subgroups within the school fail to meet certain proficiency *levels*.

Finally, Kane warns that using the NAEP tests to study the impact of state accountability standards is problematic for the 1990s because the reported scores exclude students whose disabilities resulted in accommodations while taking the test, such as extra time or having test questions read to them. The proportion of students granted accommodations increased after the Individuals with Disabilities Act of 1996, and Kane cites the cases of two states with prominent accountability systems that also had large increases in exclusion rates.

### DO STUDENT ACHIEVEMENT STANDARDS RAISE PERFORMANCE?

The majority of U.S. states now have or are phasing in examinations that students must pass in order to graduate from high school. Examples mentioned at the conference include the MCAS (Massachusetts Comprehensive Assessment System) and the FCAT (Florida Comprehensive Assessment Test). **John Bishop**'s paper analyzes the likely effects of these new exams, based on evidence from longer-standing testing programs that he calls curriculum-based external exit exam (CBEEE) systems. These examinations evaluate students' mastery of the high school curriculum, and individual-student scores play a role in determining university admission. In the United States, the primary example is the Regents exam system in New York. Such examinations are more prevalent in other countries.

Where they are found, both the new graduation tests in the United States and the longer-standing CBEEEs cover all or almost all high school students, define achievement relative to an absolute standard, vary according to the curriculum in a specific geographical area (such as a state, province, or country), and are controlled by the same education authority that designs the curriculum and funds elementary and secondary education. In addition, both types of examinations have consequences for students and schools, although CBEEEs generally have been more oriented toward measuring student achievement rather than determining who graduates.

In five separate samples, Bishop finds that the existence of CBEEE systems improves academic performance substantially—one-half to twothirds of a grade-level equivalent. He measures performance according to scores on widely applied standardized tests that are not curriculumbased and, using regression analysis, compares these scores in areas with and without CBEEEs. The five test measures are the national average performance of eighth graders on the Third International Math and Science Study (TIMSS); the achievement of 14-year-olds in the reading literacy study of the International Association for the Evaluation of Educational Achievement (IEA); the national average performance of 15-year-olds in the Program for International Student Assessment (PISA); the Canadian province average performance of 13-year-olds on the International Assessment of Educational Progress (IAEP); and New York versus other states' average high school student performance on the Scholastic Achievement Test (SAT).

Bishop hypothesizes that CBEEEs increase achievement through various positive incentives for students, parents, teachers, and school administrators. He argues that such tests provide an offset to the problem of peer pressure against studying. Teachers no longer act as judges of their students' performance, but in effect become coaches for their students, helping them pass exams that are established by an authority outside the classroom. The Canadian study was supplemented by additional data showing that schools in provinces with CBEEEs scheduled more hours of math and science instruction and assigned more homework. Students in these provinces reported spending more time reading for pleasure and devoted a greater share of television-viewing time to educational programs.

While Bishop expects the new high school exams in the United States to have some of the same effects as CBEEEs, he notes some crucial differences. The new exams set minimum competencies for graduation. Thus, if anything, they are likely to result in more class time being devoted to practicing low-level skills, as opposed to inducing teachers to spend more time on cognitively demanding skills. Furthermore, if only a pass–fail signal is generated by the exams, and if passing is necessary to graduate, Bishop argues that standards are likely to be set low enough to allow almost everyone to pass the test after multiple tries. Thus, these exams are not likely to spur the great bulk of students to increase their effort.

In commenting on Bishop's paper, **David Figlio** expresses support for the view that higher standards can improve student performance. He notes that Bishop's findings are complementary to other research that finds that students learn more and behave better when they have a teacher with high grading standards. However, Figlio is skeptical that the introduction of comprehensive, curriculum-based tests can increase student performance as much as Bishop finds.

Figlio argues that some reverse causality is at work in Bishop's study. For example, the fact that children in provinces with CBEEEs read more for pleasure and devote a greater share of their television-viewing time to educational programs may be attributes of their communities rather than outcomes of the CBEEEs. Parents in provinces imposing CBEEEs are likely to have a greater preference for certain types of instruction than parents living elsewhere, and these tastes result in a difference both in curricular emphasis and in setting standards. To attribute the full difference in test scores to the CBEEEs, and none to parental preferences, overstates the role of these exams.

Figlio calls for more research on the distributional consequences of testing. For example, although Bishop argues that CBEEEs induce students to work harder, it is also plausible that they may discourage low achievers, causing them to drop out in greater numbers.

Finally, Figlio expresses concern about the simultaneous existence of school standards and student standards. The No Child Left Behind Act removes federal education aid from "failing" schools. The threat of such

a penalty may inhibit states from identifying schools as poor performers. Figlio suggests that the federal legislation may have been a factor in causing Florida to delay implementation of higher standards for its comprehensive assessment test. Citing separate research showing that student performance is lower in schools that give merit pay to all or most teachers (regardless of individual teacher productivity) than it is in schools with no merit pay, Figlio speculates that imposing low student standards might be worse than having none at all.

Ellen Guiney warns that educational reformers need to create a coherent system in order to improve instruction in urban classrooms, where students tend to exhibit the greatest learning deficiencies. As currently implemented, standards-based reforms rest on assumptions that do not hold in large school districts. Virtually no large district provides timely information to school principals and teachers about what individual students are and are not learning. Timeliness is particularly critical in the urban environment, since students change schools frequently. Furthermore, low-achieving students have little confidence that schooling has value for them, since their own experience is largely contradictory.

Guiney argues that teachers often do not know how to assess individual student progress or how to design an appropriate course of study based on individual need. This problem is acute in urban schools, where teachers fear losing control of the classroom and, therefore, engage in little discourse with their students.

Teachers are not evaluated on the basis of how their students perform on curriculum-based exit examinations, which weakens the incentive to improve instruction. Moreover, even if incumbent teachers were found to be poor performers, a supply of other, well-prepared instructors ready and willing to step into urban education does not exist. Finally, urban schools and school systems lack information on how to organize financial and human resources so as to improve instruction. More research, and more dissemination of such research, are necessary.

### POLICY IMPLICATIONS: A PANEL DISCUSSION

The concluding panel focused on policies to improve educational outcomes. **Chester Finn** outlines four "theories of action" that have driven educational reforms and assesses their relative strengths and weaknesses. The first two approaches operate chiefly within the framework of familiar institutional arrangements. One theory is that school authorities are committed to improving educational outcomes and have the expertise to do so. The appropriate action in this case would be to provide additional resources to the existing educational system. This would likely lead to changes such as smaller class sizes, longer school days, introduction of new textbooks, and added use of technology. Finn believes such a policy works only in the case of unusually high-quality leadership within the school system. A second theory holds that school officials are motivated to improve but need further training on effective organization and teaching methods. The policy response in this case would be added involvement of outside education experts. Finn views independent professionals as a useful adjunct to educational reform, but he does not believe they can be entrusted with the responsibility of making reform happen.

Finn's remaining two theories of action view outsiders as the drivers of education reform. Government-driven reform is premised on a greater need for higher levels of government to be involved in elementary and secondary education. For example, the No Child Left Behind Act calls for state governments, backed by the federal government, to develop educational standards, test performance against these standards, and institute a set of incentives to ensure positive results. An alternative view is that educational reform should harness the power of market forces by introducing competition among schools and providing families a choice of schools. Unlike the other theories of action, Finn notes that marketdriven reforms have not yet been tried on a large scale.

Finn argues that government-driven and market-driven reforms are useful complements. The market-driven approach, by itself, lacks informed consumers. This problem can be obviated by the introduction of government standards and testing. On the other hand, while a government-driven accountability system is good at identifying failing schools, Finn argues that market-oriented alternatives (such as charter schools) are much more effective in implementing corrective actions.

The second panelist, **Alan Merten**, emphasizes that some of the same forces influencing reform in primary and secondary schools are affecting higher education. As a result of broader access to post-secondary education, the typical university student in the United States is no longer between the ages of 18 and 22 years, enrolled full time, and living on campus. Therefore, the structure of learning must be reformed. Courses of fixed duration with grades from failing to excellent make less sense than before. Thus, universities are beginning to adopt the model that the "time and place" of learning are variable, but that minimum standards must be set for knowledge gains. In addition, Merten argues that educational leaders at the post-secondary level must become more willing to take risks, measure the relevant outcomes, become more effective managers of resources, and learn from failures.

Merten observes that, in their quest for accountability and costcutting, public officials have become less supportive of education. He urges education leaders at all levels to become more aggressive, not only in managing resources more efficiently but in making the case for the allocation of more adequate resources for education. This requires clarifying the link between education and economic and social prosperity. He notes that the U.S. educational system has been instrumental in expanding opportunities for women, ethnic minorities, and non-U.S. citizens. Further progress is needed in light of the continuing need to develop a workforce for the information economy. Unfortunately, Merten notes, the terrorist attacks of September 11, 2001, have engendered some moves to restrict access of foreigners to U.S. higher education institutions.

Finally, Merten lists three features that distinguish higher education markets in the United States. These are intense competition among providers of education, merit-based pay, and compensation that differs according to academic discipline and area of expertise. Given the excellent worldwide reputation of this nation's universities, Merten urges public policymakers to assess whether these structures may usefully be adopted at lower levels of education.

Panelist **Richard Murnane** argues that unless public officials, teachers' unions, business groups, and the community at large band together to support reform of urban public schools, we are likely to see everincreasing diversion of public resources to alternative schools, with the possible demise of public schools "as we have known them." As a first step, effective school reform must encompass the development of measures of student outcomes that are meaningful in the context of current labor markets. To earn a decent living, workers must increasingly engage in nonroutine problem solving and in communicating the meaning of information. Tests that are geared only toward measuring students' reading comprehension and their ability to perform computations are not adequate in achieving effective reform. Only selected state testing programs currently go beyond these outdated standards.

Murnane argues that school reform must also encompass efficient analysis of individual student performance and the training of teachers to improve student outcomes. Otherwise, the information provided, even in good testing programs, will not be put to its desired use. Beyond such reforms within the traditional schooling context, Murnane calls for added resources to support summer learning programs for low-income children, so as to prevent them from falling behind their higher-income peers during the periods when they are not enrolled in school.

Murnane draws lessons from the experiences to date with alternative schools. Charter and voucher schools have been reluctant to accept students with disabilities, students whose first language is not English, disruptive students, and students who change schools frequently. Policymakers should interpret their reluctance as evidence that current funding formulas do not compensate schools adequately for educating these categories of students. Murnane urges the creation of a level playing field on which traditional public schools compete with charter schools and voucher schools.

Experiments with alternative schools also offer examples of resource use that could be applied in traditional public school settings if certain
institutional rules were made more flexible. As an example, Murnane cites Boston's experience with pilot schools, which are staffed by personnel who have agreed to waive certain elements of teachers'-union work rules in exchange for greater flexibility in designing and implementing instructional programs. That these schools are attracting talented teachers and pursuing innovative educational programs is a testament to their success.

The final panelist, **Warren Simmons**, notes that the No Child Left Behind legislation has set education goals for 2014 that are far more ambitious than those contained in prior versions of standards-based reforms. Like these earlier efforts, Simmons believes the current moves are doomed to fall short of their goals unless standards and assessments are integrated into the other aspects of education policy, such as professional development, curriculum development, school funding, public engagement, and school organization. He points out that these various aspects of education policy are likely to be especially uncoordinated in a federalist system like that in the United States, where federal, state, and local governments, as well as school districts and the private sector, all contribute to the provision of education.

Simmons argues that large gaps continue to exist between our current levels of educational attainment and our desired levels. The existing educational system has been effective, at most, in moving elementary and middle school students from substandard to basic levels of achievement. It has not been effective in raising children's performance to proficient levels, in making progress in high schools, or in closing the gap between white students and minorities—particularly African Americans and Hispanics.

Simmons emphasizes that instead of continuing to focus on individual school performance, reform should concentrate on systemic improvement in the education system. This requires developing an education leadership made up of experts from different disciplines and sectors who are committed to working toward a common agenda. Whatever is learned at a national level must be customized for local school districts by local organizations. The local efforts must involve outside agencies and organizations that are effective in communicating to the public about education reform. They must also involve groups such as social-services providers and juvenile-justice officials who deal with related issues. In the general discussion period, Simmons gave examples of ways in which state evaluation criteria for teachers, textbook purchasing decisions, and the curriculum at a major local teacher college did not keep pace with changing educational standards set by the Philadelphia school system, hampering their successful implementation.

### Address

## American Leadership in the Human Capital Century: Have the Virtues of the Past Become the Vices of the Present?

### Claudia Goldin\*

The twentieth century became the human capital century.<sup>1</sup> No nation today—no matter how poor—can afford *not* to educate its youth at the secondary school level and beyond. But at the start of the twentieth century even the world's richest countries—richer in per capita terms than many poor nations are today—had not yet begun the transition to mass secondary school education. There was one exception, the nation that led the world in mass secondary and mass higher education: the United States.

The United States accomplished the feat of mass education by creating a new and unique educational pattern or gauge—I will call it a "template"—that broke from the templates of Europe. The U.S. template was shaped by egalitarian institutions—a commitment to equality of opportunity; by New World factor endowments—lots of land relative to labor; and by republican ideology—meaning democracy and pluralism.

For much of the twentieth century, the template was synonymous with a set of "virtues." That is, the template consisted of characteristics that were virtuous. Among the virtues of mass secondary education were that it was publicly funded; managed by numerous small, fiscally independent districts; open and forgiving; academic yet practical in its curriculum; secular in control; and gender-neutral in its admission. I call these characteristics virtues because they promoted and furthered mass education and thereby increased social mobility and enhanced economic growth.

What brought about the human capital century? Why and how did

<sup>\*</sup>Henry Lee Professor of Economics, Harvard University.

<sup>&</sup>lt;sup>1</sup> This address draws on Goldin (2001).



Figure 1

Note: The horizontal axis is expressed in natural log units. The markers refer to geographic groupings established by the United Nations Organization for Education, Science and Culture (UNESCO): 1 = Africa, 2 = Central and North America and the Caribbean, 3 = Asia, including the non-African Middle East, 4 = Europe, 5 = Oceania, and 6 = South America. Gross enrollment rate (GER) gives the fraction of individuals in an age group who are currently enrolled in a particular level of school (e.g., upper secondary). The age group in the UNESCO data is supposed to be tailored to those included in the school level for the particular nation. The GER can exceed 100 percent if some youths outside the usual age group, for various reasons, are enrolled in that level of schooling.

Source: Gross secondary enrollment rates: United Nations Organization for Education, Science and Culture (UNESCO) < http://unescostat.unesco.org/en/stats/stats0.htm>. GDP per capita: Penn World Tables <a href="http://pwt.econ.upenn.edu/">http://pwt.econ.upenn.edu/</a>> through the efforts of Alan Heston and Robert Summers. For more details, see appendix to Goldin (2001).

the United States lead the world in mass education for much of the twentieth century? What does this history mean for the future of education in the United States?

### Why Do I Claim That the Twentieth Century Was the Human **Capital Century?**

Even poor countries today have a far greater rate of secondary school enrollment than did the rich countries of the past. Consider Figure 1, for which the horizontal axis is real per capita income in 1990 (as represented by GDP) and the vertical axis is the enrollment rate of youths in upper secondary school in 1990. The lowest of the four stars in the figure

represents the real per capita income and the high school enrollment rate in the United States in 1900, just before secondary school education took off in the United States' high school movement.

Two quadrants in the diagram have unambiguous interpretations the northwest and southeast. I term the northwest quadrant the "good education" quadrant and the southeast quadrant the "bad education" quadrant. By the "good education" quadrant, I mean that nations found in it had lower real incomes in 1990 than the United States did in 1900 but a higher enrollment rate in 1990 than the United States had in 1900. By the "bad education" quadrant, I mean that the nations located in it had a higher income but a lower enrollment rate than the United States in 1900. No nation is in the bad quadrant and many are in the good quadrant. One can do the same thought experiment for other years. Figure 1 also contains a data point for the United States in 1920 and, once again, no country is located in the bad quadrant. The data point for 1940 places just a few countries in the bad quadrant. Only when the 1960 data point for the United States is considered do more than a handful of countries fall into the bad quadrant.

Two highly useful facts are embedded in these data and the thought experiment. The first fact—and it will be clearer in a moment—is that secondary schooling "took off" in the United States from around 1910 to 1940. The second fact is that the bad quadrant was virtually empty until the United States achieved very high enrollment rates, and the good quadrant was often brimming with countries. This demonstration suggests that even poor nations and poor people today invest in secondary schooling to a far greater degree than did the educational leader of the past. Thus, the twentieth century became the human capital century. Nations can no longer afford to be left behind in educating their people because today's technologies are produced by higher-education countries and are designed for an educated labor force.

The notions that "people skills" matter, that the wealth of a nation is embodied in its people, and that only an educated people can adopt, adapt, and innovate new technologies were voiced in America at the dawn of the twentieth century. In 1906, the governor of Massachusetts appointed a commission to study technical education and assigned the chairmanship to Carroll Wright—one of the greatest U.S. labor statisticians of all time, the first Massachusetts Commissioner of Labor, and the first Commissioner of the federal Bureau of Labor Statistics. The report of the Wright Commission concluded: "We know that the only assets of Massachusetts are its climate and its skilled labor" (Roman 1915). (Give the author half credit.) The modern concept of the wealth of nations had emerged. What mattered was capital embodied in people—*human capital*.

## Why Did the Twentieth Century Become the Human Capital Century?

In the nineteenth century, machines and natural resources, not people, mattered to the industrial giants—Britain, Germany, France, and the United States. But in the early 1900s, attention began to shift to the education of the people at the secondary and higher levels.

A new economy—as it was termed by contemporaries—had emerged in the early twentieth century. It involved a greater use of science by industry, a proliferation of academic disciplines, a series of critical inventions and their diffusion (for example, small electric motors, the internal combustion engine, the airplane, various chemical processes), the rise of big business, and the growth of retailing. A host of demandside factors increased the relative demand for educated labor and enhanced the returns to education and training.

These changes did more than increase the demand for a small cadre of scientists and engineers. They increased the demand for skilled and educated labor among the mass of workers. Firms began to seek employees with a host of general skills. They sought a white-collar and clerical staff capable of using the latest office machinery, with modern office skills (such as stenography and typing), polished grammar, and some mathematical prowess. They also sought blue-collar workers who could decipher manuals, who could use algebra, and who had a mastery of mechanical drawing and a familiarity with chemical and electrical fundamentals.

A remarkable notion had emerged around 1900—it was that schooling could make the ordinary office clerk, the shop-floor worker, and even the farmer more productive. The odd thing is that even though most industrial nations acknowledged the change from physical capital to human capital, only one did much about it until well into the twentieth century.

### How Did the United States Lead the World in Mass Education?

The demand for educated labor increased, and almost nationwide there was an outpouring of public and primarily local resources to build and staff high schools. These schools were academic (not industrial), free, secular, gender-neutral, open, and forgiving. The educational change was known then as the "high school movement." In the United States as a whole, the enrollment rate for youths in all secondary schools—public high schools, private secular and religious high schools, and the preparatory departments of colleges and universities—soared from 1910 to



Figure 2 Public and Private Secondary School Enrollment and Graduation Rates: United States, 1890 to 1970

1940, as seen in Figure 2.<sup>2</sup> In 1910, just one American youth in ten was a high school graduate, but in 1940 the median youth had a high school diploma. The contemporaneous graduation rate, expressed as a fraction of the relevant age group, also increased substantially during the same period. It is no wonder that those who lived through the early part of the period termed the change "one of the most remarkable educational movements of modern times" (California Department of Public Instruction 1914).

The high school movement was not just an urban phenomenon, and it was not just a New England phenomenon, although it began there. It quickly spread from New England towns to the rich agricultural areas in the central part of the country and to the western states. Because the southern states had lower levels of educational attainment for much of the twentieth century and because the high school movement diffused slowly throughout the South, the national data in Figure 2 give a

<sup>&</sup>lt;sup>2</sup> High school or secondary school is historically defined in the United States as grades 9 through 12 (even if grade 9 is offered in a junior high school) and it generally includes youths from ages 14 or 15 to 17 or 18. For further details, see Goldin (1998, 1999).



Note: Graduates include both males and females in public and private schools (including the preparatory departments of colleges and universities). The number of graduates is divided by the approximate number of 17-year-olds in the region. Constant growth rate interpolations of population data are made between census years.

somewhat misleading impression of the speed of the high school movement throughout the rest of the country. High schools spread considerably faster in most other regions of the country, and graduation and enrollment rates were higher, as can be seen in the graduation rates of Figure 3. Even before 1930, graduation rates for 18-year-olds in many parts of the North, Midwest, and West exceeded the 50 percent mark.

In 1910, when the data on graduation rates begin, New England was the leading region. But by the mid-1910s, the rich states of the Pacific had closed in on New England, and by the 1920s, even the sparsely settled and agricultural states of the West North Central (consisting of states such as Iowa, Kansas, and Nebraska) had exceeded the rates achieved in New England. Only the Middle Atlantic states were left behind, but they caught up during the massive unemployment of the Great Depression, when jobs for teens evaporated overnight and education became a more attractive alternative. In 1940, as the world braced for yet another war, America could boast the most educated workforce in the world. It accomplished this feat even though, for much of the period, it had opened its doors to the poor of the world. America's success in mass secondary education resulted from its educational template and the associated virtues.

Source: State-level high school graduation data from various sources. See Goldin (1998).





Note: The data refer to the number of youths in public and private upper and lower secondary schools (of the types listed) ranging from those who turned 15 years old during the school year to those who turned 19 years old during that year. Thus, the age group under consideration is approximately all 15- to 18-year-olds, plus one-half of 14- and 19-year-olds. No youths in elementary schools or colleges and universities are included even if they were in the included ages. The procedure ensures consistency but implicitly favors countries, such as the Nordic nations, that have late starting ages.

The computation for the United States assumes 100 percent enrollment for the 14-year-olds and then adds all those enrolled in ninth through twelfth grades and divides by the number of youths aged 15 to 19.

Abbreviations are: Austria (AUT), Netherlands (NLD), Luxembourg (LUX), Germany (DEU), Belgium (BEL), Switzerland (CHE), France (FRA), Ireland (IRL), Great Britain (GBR), Denmark (DNK), Norway (NOR), and Sweden (SWE). Countries are listed in order of the full-time general secondary school enrollment rate.

Source: For European nations, Goldin (2001), Figure 1. For United States, U.S. Department of Education (1993), Tables 1 and 9.

In contrast to the U.S. template, European templates were characterized by quasi-public or private funding and provision, by the high standards of an unforgiving system, by the unity of church and state, and by a "boys come first" attitude. The German, British, and French templates or systems, while different in their details, had much in common—strict standards, individual accountability, severe tracking at early ages, and higher education for a small, elite corps. Most of these systems had centralized bureaucracies and finances, and some had elaborate apprenticeship systems.

By the mid-1950s, the United States' lead in the human capital century was astoundingly large. A wide gap existed between the education of youth in Europe and in the United States. Across the 12 European countries in Figure 4, only one (Sweden) had a full-time, general education enrollment rate for 15- to 19-year-olds that exceeded 20 percent. In addition to Sweden, just two nations had a full-time general plus full-time technical educational enrollment rate that exceeded 30 percent. The U.S. enrollment rate for the same age group in 1955 was almost 80 percent. Even if one adds to the European data youths in part-time technical education, enrollment rates would still be considerably lower than in the United States. Only in the past three decades has the difference between the secondary school enrollment rate of the United States and that of Europe been largely eliminated and the lower quality of U.S. secondary school education become a major U.S. domestic issue.

Why did the United States at the turn of the twentieth century break from the educational and training templates of Europe and pioneer a novel form of secondary education? Why did Europeans believe that Americans were wasting resources by educating their masses? Why did Americans reject a highly specific, on-the-job, industrial form of education (such as the British, Danish, and German apprenticeship systems) in favor of one that was general, school-based, and academic? The answers to these questions concern basic differences between the New World and the Old World.

Formal, general education is more valued when geographic mobility and technical change are greater. School, not an apprenticeship and job training, enables a youth to change occupations over his lifetime, to garner skills different from his parents', and to respond rapidly to technological change. The U.S. template was not wasteful in the technologically dynamic, socially open, and geographically mobile New World setting. And, more important, it probably enhanced the dynamism.

Follow my reasoning thus far: A host of changes beginning in the late nineteenth century increased the demand for certain skills and knowledge. A set of republican institutions enabled the United States to respond to the increased demand for skill; these institutions, together with a set of New World preconditions (such as a high ratio of land to labor), meant that the United States responded to the technological imperative in a particular way. By the early twentieth century, the United States began to endow a large fraction of its youth with skills in formal, school-based, academic settings, using the U.S. template. The United States achieved mass secondary (and later mass higher) education because of a set of virtues that enabled the supply-side institutions to respond to the demand-side shift.

How did the virtues accomplish so much? Take decentralization, for example. In a state where public support for school expansion was less than 50 percent, the existence of numerous small, fiscally independent districts would enable high schools to diffuse. People choose where to live, and small districts are generally more homogeneous than are large districts with respect to income, ethnicity, religion, and cultural values. It is likely, therefore, that individual preferences for public goods are more similar the smaller the geographic area. Greater homogeneity means that the public good—education in this case—might get funded by some of the districts, whereas there would be no funding if the district were the size of the state. In contrast to the United States, educational decisions were highly centralized in much of Europe. National legislation (in Britain and France, for example) was required to fund secondary school expansion, and it initially diffused more slowly than it did in the United States.

In the United States, about 130,000 separate school districts existed around 1925, but many were tiny common school districts of the open country, and some did not have the ability to set their own tax rates. That still left tens of thousands of fiscally independent school districts of a large enough size in the early part of the twentieth century to establish a public secondary school. These relatively small, fiscally independent school districts implicitly competed with each other to attract residents.

In work that Lawrence Katz and I have done using archival records from a unique state census, we found that an additional year of high school at the start of the high school movement in 1915 added more than 12 percent to the earnings of young men (18 to 34 years old). This return was almost double that for an additional year of secondary school in 1955.<sup>3</sup> Returns were substantial even within various occupations. That is, whether a youth were somehow destined to be a blue-collar or a white-collar worker, there would still be significant returns to further education. The return to education, furthermore, was as high for farmers as it was for those in nonagricultural occupations.

What impact did the U.S. template have on economic growth and individual welfare? I'll give just one part of the answer: It had a major impact on economic inequality.<sup>4</sup> As more individuals gained more years of education in the first half of the twentieth century, inequality declined. The structure of wages narrowed, wage ratios for higher-skilled relative to lesser-skilled positions fell, and the returns to education decreased. All of the data sets I have examined show declining inequality for the period from the late 1910s to the 1950s. And they also show rising inequality after the mid-1970s. If we think of the wage structure as being the result of a race between technology and education, then education ran faster than technology in the first half of the century, and technology ran faster than education in the second half. Interestingly, technology does not appear to have accelerated after the 1970s. Rather, advances in educational attainment slowed down, in part because of demographics. But that issue must wait for another talk.

<sup>&</sup>lt;sup>3</sup> See Goldin and Katz (2000).

<sup>&</sup>lt;sup>4</sup> For evidence on changes in inequality across the twentieth century, see Goldin and Katz (2001).

#### Have the Virtues of the Past Become the Vices of the Present?

The U.S. template (characterized by virtues) succeeded during the first half of the twentieth century, and for some time after, it did better than those of other nations. The system produced far more educated citizens and workers. It did not, by and large, reinforce class distinctions but, rather, it enabled economic and geographic mobility and resulted in a large decrease in inequality in economic outcomes. It may also have increased technological change and thus labor productivity, although that is far more difficult to prove.

The virtues I have mentioned include the following: education that was publicly funded and publicly provided; an open and forgiving system; an academic yet practical curriculum; numerous small, fiscally independent school districts; and secular (not church) control of schools. But these characteristics are no longer seen as uniformly virtuous. To some, they now constrain, rather than further, education. For example:

- Public or community funding and public provision were the hallmarks of the common school system. But vouchers— public funding but private provision—and charter schools are now being used and considered for use to increase competition. (Thomas Downes discusses these subjects in his paper for this conference.)
- An open and forgiving system without tracking at early ages was seen as egalitarian and non-elitist. But this type of system is now viewed as lacking both standards and accountability. Almost all states today have standards for grade promotion, high school graduation, school funding, and teacher retention. Some of these standards are strict and have serious consequences for those who do not pass. (Eric Hanushek and Margaret Raymond, and John Bishop, in their contributions for this conference, assess whether standards and accountability have positive effects on a variety of outcomes and, therefore, whether they are truly virtuous.)
- A general, academic education for all may enhance flexibility *ex ante*, but may, *ex post*, leave many behind and may have worsened rising inequality. Some have recently espoused technical and vocational training for certain youths.
- Although a decentralized system of small, fiscally independent districts competing for residents once fostered educational investments, these systems are now seen as producing serious funding inequities. State equalization plans are currently in effect in most states, although some plans (such as that in California) have led many to exit the public system and may actually reduce spending per child in poor districts.

(Thomas Downes summarizes this literature in his paper for this conference.)

• The separation of church and state encouraged a common education for all. But an insistence on the secular control of public funds would mean that Catholic and other churchbased schools could not receive publicly funded vouchers, even in academically failing school districts where other private schools are unavailable to poor students. The recent Supreme Court ruling on this important issue (*Zelman* v. *Simmons Harris* on June 27, 2002) may widen the use of vouchers by denominational schools, not just by those in failing school districts.

In conclusion, the twentieth century was the human capital century. America led other nations by a wide margin in the provision of general, formal education to the masses and did so because of characteristics virtues—that were shaped by New World endowments and republican ideology. Almost all of these virtues are now being questioned, and in the twenty-first century an entirely new set of virtues could emerge.

#### References

- California Department of Public Instruction. 1914. 1913/14 Biennial Report of the Superintendent of Public Instruction. Sacramento, CA: The Superintendent.
- Goldin, Claudia. 1998. "America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century." *Journal of Economic History* 58 (2): 345-74.
- ——. 1999. "Egalitarianism and the Returns to Education During the Great Transformation of American Education." *Journal of Political Economy* 107 (6): S65-94.
- 2001. "The Human-Capital Century and American Leadership: Virtues of the Past." Journal of Economic History 61 (2): 263-92.
- Goldin, Claudia and Lawrence F. Katz. 2000. "Education and Income in the Early 20<sup>th</sup> Century: Evidence from the Prairies." *Journal of Economic History* 60 (3): 782-818.
- 2001. "Decreasing (and Then Increasing) Inequality in America: A Tale of Two Half Centuries." In *The Causes and Consequences of Increasing Inequality*, edited by F. Welch, 37-82. Chicago, IL: University of Chicago Press.
- Roman, Frederick William. 1915. Industrial and Commercial Schools of the United States and Germany: A Comparative Study. New York: G.P. Putnam's Sons.
- U.S. Department of Education. 1993. 120 Years of American Education: A Statistical Portrait. Washington, DC: Government Printing Office.

## EDUCATIONAL ATTAINMENT AS A CONSTRAINT ON ECONOMIC GROWTH AND SOCIAL PROGRESS

### Yolanda K. Kodrzycki\*

Perceptions of the economic problems posed by inadequate educational attainment in the United States have changed over time. During the first part of the past half-century, U.S. educational reforms were driven heavily by political and economic competition with other parts of the world. The National Defense Education Act was passed in 1958 in response to the successful launch of the Soviet Sputnik. This legislation articulated the Cold War education challenge as the need to "develop as rapidly as possible those skills essential to the national security" (Title 1, A). In 1983, the National Commission on Excellence in Education, formed at the behest of the U.S. Secretary of Education, issued its findings on the quality of the education system in *A Nation at Risk*. The report warned of rival nations matching or even surpassing U.S. educational levels and saw the manifestation of a decline in U.S. productivity growth "as one great American industry after another falls to world competition."

In the last several decades of the twentieth century, the focus of national education policy shifted gradually from achieving international prowess to making progress on economic and social equality within the United States. The first major linkage of national education reform to equity concerns came in the Elementary and Secondary Education Act of

<sup>\*</sup>Assistant Vice President and Economist, Federal Reserve Bank of Boston. The author thanks Katharine Bradbury for generously sharing her insights and computations concerning educational attainment and earnings. Lynn Browne provided valuable guidance on an early draft. Additional colleagues from the Federal Reserve Bank of Boston and the conference attendees offered many perceptive comments, some of which have been taken into account in this final version, and Stephan Thernstrom pointed out a data error that has since been corrected. Mary Fitzgerald provided excellent and extensive research assistance in all phases of preparing this paper, and Krista Becker helped in obtaining and organizing reference materials.

1965, which focused on the needs of low-income children as part of the overall "War on Poverty." With the disintegration of the Soviet Union and the generally good U.S. economic performance in the second half of the 1980s and throughout the 1990s, educational reforms became even further disassociated from the language of international conflict and competition. Although it appeared solid on the whole, America's economic growth offered differential benefits to different groups, as workers with high educational attainment increasingly gained access to relatively higher-paying jobs, while real pay for workers with low educational attainment decreased over time. Thus, the Goals 2000: Educate America Act, passed with bipartisan support in 1994, focused on the problems associated with continuing educational achievement gaps among racial groups and between persons who were proficient in the English language and those who were not. Equalizing opportunity within the United States remained the primary goal behind the landmark No Child Left Behind Act of 2001.

In recent years, technology's importance in economic growth and the need to educate and train a technologically oriented workforce have been increasingly emphasized. Although earlier literature, most notably Richard Freeman's *The Overeducated American* (1976), warned of periodic gluts of college graduates as cycles of labor supply and labor demand did not coincide with one another, recent studies such as the National Research Council's *Building a Workforce for the Information Economy* (2001) have been more inclined to see tightness in scientific and technical fields as a secular feature of the economy.

This paper investigates the evidence behind these shifting perceptions of the educational problem. It starts by reviewing the changes in overall educational attainment in the United States during the past several decades and by analyzing the implications for past and future economic growth. The paper then examines the educational attainment of different demographic groups in the population and the ramifications for social progress. Finally, the paper addresses arguments about mismatches in the supply of and demand for technically trained workers.

The paper reaches two broad conclusions. First, a growing body of evidence indicates that improving the quality of U.S. education, both on average and for specific population groups, should be of more concern than increasing the quantity of schooling. Second, as minority racial and ethnic groups account for a growing share of the U.S. population, improving their educational opportunities goes hand in hand with overall economic growth objectives.

### THE FACTS ON OVERALL EDUCATIONAL ATTAINMENT

Educational attainment in the United States has changed over the past 30 years. This section discusses whether or not overall educational



attainment is increasing and reviews the educational rankings of the United States compared to other countries. In examining these issues, the paper uses alternative measures of educational attainment, including both the extent of schooling (quantity) and the amount of knowledge obtained (quality) during these years of schooling.

### **Overall Trends**

The U.S. population has become far more schooled during the past three decades. The share of the population 25 years and over who have completed high school rose from 55.5 percent in 1970 to 84.0 percent in 2000. The share completing four years of college rose from just 11.0 percent to 25.5 percent during this period (Figure 1).

However, much of the increase in schooling since the 1970s is due to the dying out of older generations with comparatively little education, rather than steadily growing educational attainment among younger generations. Individuals who are currently 25 to 29 years old have very similar educational attainment to their predecessors' levels of two decades ago. The share of 25- to 29-year-olds that completed high school increased from about 76 percent in 1970 to 85 percent in 1977 (Figure 1). This percentage stayed virtually constant until 1991 when it began increasing slightly, reaching 88 percent in 2000. Similarly, college completion rates among 25- to 29-year-olds increased in the 1970s but then held steady at around 25 percent throughout the 1980s and early 1990s.



Figure 2 NAEP Long-Term Trend Average Test Scores for 17-Year-Olds

College completion rates began to rise again in the second half of the 1990s, reaching about 29 percent by 2000.

While the number of years of schooling provides a rough estimate of the educational levels of the population, examining the knowledge gained during these years provides a useful measure of the quality of educational attainment. Murnane and Levy (1996) identified three categories of basic skills that are increasingly demanded by U.S. employers and that are necessary to earn at least a middle-class income in the United States. The first category includes hard skills such as mathematics, problem-solving, and reading ability. Relying on standardized test scores for these data, the most consistent time series comes from the National Assessment of Educational Progress (NAEP) long-term trend tests. Versions of these tests have been administered to nationally representative samples of 9-, 13-, and 17-year-olds periodically since 1969.1 The NAEP and other nationwide tests measuring educational achievement trends do not assess the remaining two categories of skill sets identified by Murnane and Levy: "soft" skills such as the ability to work in groups and to make effective oral and written presentations, and the ability to use

<sup>&</sup>lt;sup>1</sup> The first NAEP long-term trend test in science was administered in 1969. However, the early administrations of this exam are not reliably comparable to later tests because of changes in the questions and methodology. A similar problem exists in mathematics. In order to ensure consistency, only the test scores of the assessments beginning in 1977 for science and 1978 for math were examined.



Figure 3 Average Years of Schooling for Populations Aged 25 and Over, U.S., World, and Regions of the World

Note: Barro and Lee did not make adjustments in the U.S. data to compensate for definitional changes in the CPS from 1992 onwards. Source: Barro and Lee (2000). Individual country data provided by authors through the Center for International Development at Harvard University.

personal computers to carry out simple tasks such as word processing. (Additional information on access to computers is presented later in this paper.)

Figure 2 shows the average NAEP long-term trend test scores of 17-year-olds. Math, science, and reading scores all increased during the 1980s after decreasing in the 1970s. However, the 1990s saw smaller increases and showed some indications that 17-year-olds are losing ground. As of 1999, math scores were only slightly above their previous peak in 1992. Science scores continued to increase in the early 1990s but have retreated since 1996. Reading scores reached a plateau in the late 1980s and early 1990s, fell in the mid-1990s, and are currently holding steady at these lower levels.

## International Comparisons: Wide Disparities across Countries and a Mixed Record for the United States

The United States leads the world in the average amount of education received by the population, with 12.2 years of schooling in 2000 (Figure 3). More generally, there exist large and persistent disparities in the educational attainment between advanced and transitional economies and developing economies. In 2000, the population of advanced and





Note: The indicated sample sizes include the United States. Source: Lee and Barro (1997). Data provided by authors through the World Bank web site, www.worldbank.org.

transitional nations on average had almost 10 years of schooling, while those in developing nations had less than five years. The overall gap has changed very little since 1960. Among the nations classified initially as developing, East Asian and Pacific countries have made substantial improvement over the past 40 years. They lead the developing nations in educational attainment, averaging 6.5 years, which is close to the worldwide average. In contrast, sub-Saharan Africa is at the bottom of the developing countries, with average years of schooling less than four years.

Despite the U.S. lead in having a highly schooled population, the U.S. educational system has not outshone the rest of the world in terms of student achievement at given levels of education. Figure 4 shows U.S. and other countries' scores on five international mathematics tests administered to 13- and 14-year-olds between 1964 and 1998. For each year, the figure shows the U.S. average score compared to advanced and transitional countries on the one hand and developing countries on the other.<sup>2</sup> The different years are not strictly comparable, as the methodologies, the groups of participating countries, and the coverage within

<sup>&</sup>lt;sup>2</sup> Countries are defined as being in the same categories as in Figure 3.

countries for each test have all varied. Nevertheless, U.S. teens consistently place towards the lower to middle end of nations tested in mathematics, including those nations classified as developing.<sup>3</sup> The general picture for science, not shown, is similar.

# The Role of Educational Attainment in Economic Growth

The previous section indicated that the U.S. population has increasingly received more years of schooling, but that the gains have slowed as progress among younger cohorts has diminished. Furthermore, the quality of education in the United States, through high school, is not impressive and is no longer improving for the average student—at least as measured by standardized tests focusing on reading, mathematics, and science. This section explores the impact of educational attainment on U.S. growth, both historically and in the future.

### Causes of Growth in the United States

The most detailed accounting of the role of educational attainment in U.S. growth is found in a series of papers by Dale Jorgenson and various co-authors. These studies conclude that increases in labor quality via rising educational attainment have had a measurable effect on economic growth in recent decades.

As detailed initially in Jorgenson, Gollop, and Fraumeni (1987), the studies analyze the contributions to U.S. economic growth from capital and labor inputs and productivity. Labor's contribution comes from both increases in work hours and increases in the quality of the workforce. In the most recent of these studies, Jorgenson, Ho, and Stiroh (2002) estimate that increases in labor quality, via a more highly educated workforce, contributed an average of 0.3 percentage point per year during the period 1958–99. Overall economic growth (value added) during this period was 3.4 percent per year, and growth in output per hour worked was 1.8 percent per year.

Of the subperiods highlighted by Jorgenson, Ho, and Stiroh, the highest contribution of labor quality was in the first half of the 1990s (0.4 percent per year), and the lowest contribution was in the second half of the 1990s (0.2 percent per year).<sup>4</sup> The reason for the drop in the most

<sup>&</sup>lt;sup>3</sup> Admittedly, developing countries are likely to administer the test to only a small fraction of 13- and 14-year-olds, since the average years of schooling in these countries is low.

<sup>&</sup>lt;sup>4</sup> A related paper by Oliner and Sichel (2000) estimated that growth in labor quality contributed 0.22 percentage point to annual economic growth from 1974 to 1990, 0.44 percentage point from 1991 to 1995, and 0.31 percentage point from 1996 to 1999. They

recent five-year period is that as the unemployment rate fell in the late 1990s, many workers with relatively less education and experience entered the ranks of the employed labor force.

As valuable as the calculations of Jorgenson and his co-authors are, they may possibly understate the overall importance of education in U.S. economic growth in recent years. The neoclassical framework used in these studies measures the contribution of education to workers' productivity, but it does not attempt to quantify the role of rising educational attainment in making capital more productive. An increase in the supply of educated workers increases the market size for technologies that are complementary to educated labor and may induce the use of such technologies (Acemoglu 1998). This relationship is illustrated by comparing recent information technologies with older inventions: It takes more education to use a computer than to turn on an electric light switch or to drive an automobile. Thus, some of the growth that Jorgenson and his co-authors attributed to the greater use of information technologies (0.5 to 1 percent in the 1990s) might not have come about were it not for the education of the labor force.<sup>5</sup>

#### Projections of Stagnating Labor Force Quality

In concert with the analysis in the prior section of this paper, Ho and Jorgenson (1995) noted that the educational attainment of the 25- to 34-year-old age group has changed relatively little since the early 1980s.<sup>6</sup> Accordingly, they predicted that this relatively small increase in educational attainment will translate into gradually diminishing educational attainment increases for the workforce as a whole as these young workers account for a growing share of the overall U.S. labor force. Thus, the contribution of labor quality to growth is likely to be smaller in coming decades compared to what it was in the 1960s through the mid-1990s.

estimated the following annualized growth rates in real nonfarm business output: 3.1 percent in 1974–90, 2.8 percent in 1991–95, and 4.9 percent in 1996–99. Thus, Oliner and Sichel are in general agreement with Jorgenson, Ho, and Stiroh about the relative importance of improvements in labor quality in overall economic growth. Unlike Jorgenson, Ho, and Stiroh, Oliner and Sichel do not view the latter part of the 1990s as being a period of low growth in labor quality, but they agree that the first part of the 1990s saw higher growth in labor quality.

<sup>&</sup>lt;sup>5</sup> Similarly, Oliner and Sichel estimated that greater use of information technologies contributed 0.5 percentage point to annual economic growth from 1991 to 1995 and 1.1 percentage point from 1996 to 1999.

<sup>&</sup>lt;sup>6</sup> This is especially true at the low end of the educational distribution. In 1982, 10.3 percent of 25- to 34-year-olds in the workforce had not completed high school. This share was 9.8 percent in 1999. At the high end, the share completing four or more years of college was fairly stable at about 27 percent between 1982 and 1994, but according to Ho and Jorgenson's updated tables (1999), it increased another 4 percentage points between 1995 and 1999.

A separate set of projections by Ellwood (2001) underscores this point. Both the overall size of the labor force and the share with a college degree are expected to show much smaller increases between 2000 and 2020 than between 1980 and 2000. The total U.S. labor force grew from 79.8 million in 1980 to 118.5 million in 2000, nearly 50 percent. Given the age mix of the current population and reasonable assumptions about immigration, the labor force is expected to expand by no more than 19 million, or 16 percent, between 2000 and 2020. The fraction of the labor force with a college degree increased from 21.6 percent in 1980 to 30.2 percent in 2000. If subsequent cohorts have the same education at age 25 as the 25-year-old cohort of 2000, the share of the labor force with a four-year college degree would increase only to 31.7 percent by 2020. Even under optimistic assumptions about rising educational attainment, the college share would increase only to 35.2 percent.<sup>7</sup>

Despite the strong presumption that the share of the labor force that is college-educated is likely to stagnate in the next two decades, the implications for U.S. economic growth are unclear. Under one view, this would constrain growth both through slowing worker quality (as in the Jorgenson studies) and by retarding the development and dissemination of new technologies. Under another view, the mix of contributions to future growth may be different from what it has been in the past, but the high number of years of education of the current and entering workforces may be sufficient to assure undiminished growth.<sup>8</sup> The remainder of this section reviews studies that cast further light on these predictions.

### The Links between Education and Productivity

An article by Lucas (1988) set out to explicate the "mechanics of economic development" by focusing on the potential importance of human capital in enhancing the productivity of an economy's labor and physical capital. Inspired by this largely conceptual study, a series of subsequent empirical papers on "endogenous growth" investigated whether the average level of educational attainment, measured at a certain point in time, has a positive effect on a nation's per capita income growth in *subsequent* years. Some of these studies also examined whether increases in educational attainment have a contemporaneous effect on the

<sup>&</sup>lt;sup>7</sup> Ellwood's "high-growth" scenario assumes that graduation rates from high school rise 0.25 percentage point per year over the next 20 years, the entry rate from high school into some college rises by 1 percentage point per year, and the entry rate from some college to college graduation rises by 1 point per year.

<sup>&</sup>lt;sup>8</sup> For example, technological development might be redirected toward technologies that are less dependent on rising educational attainment for their adoption. Additionally, investment in physical capital might conceivably accelerate to offset slowing human capital investment.

rate of growth.<sup>9</sup> All the empirical studies conclude that there is a positive association between education and growth. However, because of measurement issues inherent in comparing countries with different educational systems and economies, disagreement continues to exist about how strongly and quickly education *causes* growth.<sup>10</sup>

On the whole, the endogenous growth literature to date has more definitive implications for developing countries than for developed countries such as the United States. Benhabib and Spiegel (1994) and Krueger and Lindahl (2001) found that countries with very low levels of educational attainment tend to grow slowly, all else equal. One explanation, supported in the former study, is that these countries lack the know-how to adopt the more productive technologies that are available elsewhere. This conclusion provides a pessimistic view of the growth prospects of the least educated of the developing countries.

Despite disagreements about the magnitude of the effect, the literature provides new evidence that quality of education may have an impact on economic growth, independent of quantity of education. Hanushek and Kimko (2000) examined the relationship between cross-country growth rates from 1960 to 1990 and average scores on various international math and science tests. In a closely related study, Barro (2001) studied per capita growth across countries in three time periods—1965– 75, 1975–85, and 1985–95—along with the students' science scores from each country. Both studies found that quality of education, as measured by standardized tests, had more explanatory power than years of schooling. (While intriguing, these studies suffer from a comparatively limited amount of international test score data, so they should not be used too literally for policy analysis.)

All in all, the empirical literature since Lucas offers guidance to the United States while stopping short of a definitive conclusion about whether future per capita income growth will slow. It suggests that future growth would be higher if the average quality of schooling were higher and if the nation continued to make progress in raising the average number of years of schooling.

<sup>&</sup>lt;sup>9</sup> This would flow out of the neoclassical growth model if labor were measured in efficiency units (as in Jorgenson's and related empirical work). It also flows out of aggregating the most commonly used microeconomic model of wage determination. The individual studies are discussed in Appendix A.

<sup>&</sup>lt;sup>10</sup> Additionally, some attempts have been made to study these issues by comparing metropolitan areas within the United States (which reduces some of the measurement problems). This within-country literature has not reached consensus about which level of education, secondary or post-secondary, matters more. See Appendix A.



Source: U.S. Census Bureau (March, various years).

# CONTINUING RACIAL AND ETHNIC DISPARITIES IN EDUCATIONAL ATTAINMENT

Perhaps more striking than the recent stagnation in educational attainment of the young, both in terms of years of schooling and knowledge acquired, is the growing gap among distinct population groups within the United States. Most notably, racial and ethnic inequalities persist in the educational attainment levels of Americans, with blacks and Hispanics continuing to be less educated, on average, than their white counterparts.

Examining the population age 25 years and older, blacks and Hispanics lag behind in both high school and college completion rates (Figure 5).<sup>11</sup> In 2000, about 85 percent of white adults had completed high school, compared to 78 percent of blacks and only 57 percent of Hispanics. The high school completion rate for "other" races—chiefly Asians—was very similar to that of whites. Twenty-six percent of white

<sup>&</sup>lt;sup>11</sup> To ensure a consistent time series going back to 1970, the data on Hispanics from the Current Population Survey include both white and black Hispanics. The "white" and "black" categories shown here include Hispanics. The category marked "other" includes Asians, American Indians, and additional races.



Figure 6 High School and College Completion Rates of U.S. Population Aged 25 to 29, by Race/Ethnicity

adults and 38 percent of the "other" racial group had completed four years of college, compared to only 17 percent of blacks and 11 percent of Hispanics. The gaps for blacks and Hispanics are not just vestiges of past social inequality; they persist among the 25- to 29-year-old age group, especially with regard to college completion (Figure 6).

The trends in black and Hispanic high school completion rates are different from one another. By 1999, black 25- to 29-year-olds had successfully reached the high school completion rates of their white cohorts (88 percent).<sup>12</sup> Hispanics have remained far behind as a result of

<sup>&</sup>lt;sup>12</sup> Past studies have shown that much of the racial gap in high school attainment has been closed by blacks via high school equivalency certificates (chiefly the GED, or General Educational Development). Differential use of the GED among racial and ethnic groups may be a source of concern if, as some studies have found, the payoff to a GED is not as high as the payoff to a regular high school diploma, exacerbating racial and ethnic inequalities. In fact, the share of GEDs among all high school finishers has risen since the 1970s, but the differences across blacks and whites have narrowed. In 1971, the number of GEDs equaled about 7 percent of all high school degrees. This share rose to about 14 percent by 1980 and hovered around 16 percent throughout the 1990s. According to Cameron and Heckman (1993), among 25-year-old males between 1979 and 1987, blacks were almost twice as likely to earn their degree via the GED than whites (13.3 percent versus 6.8 percent). However,

relatively stagnant attainment levels since the early 1980s. In 2000, 63 percent of 25- to 29-year-olds of Hispanic origin had completed high school; this was barely greater than the 61 percent rate that existed in 1982. Much of this disparity is the consequence of large influxes of relatively poorly educated Hispanic immigrants.<sup>13</sup> The U.S.-born Hispanic population in this age group shows a high school completion rate around 80 percent, much closer to the rates of their black and white peers.

Despite gains in high school degrees among the black population, black–white differences in four-year-college completion rates have not diminished over time. In fact, the black–white gap is slightly greater among young adults than in the adult population as a whole. Between 1970 and 2000, the college-completion rate among 25- to 29-year-old whites increased from about 17 percent to almost 30 percent (an increase of 13 percentage points). For young black adults, the rate increased from 7 percent to almost 18 percent (an increase of 11 percentage points). Gains in college completion rates for Hispanics over this entire 30-year period have been far more modest than both white and black gains and have been virtually nonexistent since the late 1980s.

The racial gap in educational attainment is less severe among persons who live in the suburbs, where the population has higher average educational attainment than the urban or rural populations. In 2000, 22.9 percent of black adults living in the suburbs had completed four years of college, close to the population-wide average of 25.5 percent and 5.7 percentage points below the suburban white average.<sup>14</sup> Among urban dwellers, 31.1 percent of whites had college degrees versus only 15.7 percent of blacks, for a gap of 15.4 points. Additionally, all suburban 17-year-old groups outperformed their urban and rural counterparts on NAEP tests, suggesting variance in school quality by location. The bulk of blacks continue to live in urban areas (53 percent), but an increasing share is living in suburban areas (34 percent in 2000). This location shift may possibly help raise black educational attainment in the future.

Since both young blacks and young whites have increased their rates

Current Population Survey data from 1999 indicate that among the population aged 18 to 29, 9.8 percent of blacks (males and females) received their high school degree via the GED compared to 8.6 percent of whites. Thus, while the GED has played an important role in increasing relative high school attainment levels of blacks in the past, its importance appears to have diminished over time. However, the increasing reliance on the GED for high school attainment levels is likely associated with the observed slowing effect in overall college completion rates, as those who get a GED are less likely to go on to complete higher education than those who receive a traditional high school diploma. See Boesel, Alsalam, and Smith (1998).

<sup>&</sup>lt;sup>13</sup> See Little and Triest (2002) and Clark and Jaeger (2002) for analyses of the role of Mexican immigration in the educational attainment of U.S. Hispanics.

<sup>&</sup>lt;sup>14</sup> Overall high school completion rates have become more similar over time for dwellers in urban, suburban, and rural areas, but a growing gap has appeared between rural and metropolitan populations in their shares of college-educated.

of high school and college degree completion over time, the more stagnant patterns for young adults as a whole shown in Figure 1 must be attributable to the changing composition of the U.S. population. Indeed, the total (white plus black) Hispanic share of 25- to 29-year-olds is estimated to have risen dramatically, from 5.0 percent in the early 1970s to 15.5 percent in 2000.<sup>15</sup> The relatively low educational gains of this group over time have contributed significantly to the overall stagnation for young adults. Whites' overall population share has fallen, which also serves to depress overall educational gains, but this has been partly offset by the rising share for Asian Americans. In 1970, the 25- to 29-year-old age group was 88.2 percent white, 10.6 percent black, and 1.2 percent "other." In 2000, this group was 79.7 percent white, 13.8 percent black, and 6.5 percent "other." Breakdowns of "other," available since 1989, show that the Asian-origin share is now over 5 percent.<sup>16</sup>

## Differences in Academic Achievement, Access to Information Technology, and Literacy

Knowledge assessment measures also indicate continuing racial and ethnic disparities. Differences in the white, black, and Hispanic test scores of 17-year-olds narrowed somewhat during the 1980s but many of these gains were lost in the 1990s.<sup>17</sup> The black–white gap in NAEP reading scores (Figure 7) narrowed from 52 points in 1971 to 21 points in 1988, but widened again to 31 points by 1999. The Hispanic–white gap was 41 points in 1975, 22 points in 1990, and 24 points in 1999. The basic patterns for math and science (not shown) are similar. These growing differences in the 1990s are not characterized by faster gains by white students, but rather represent declining scores among blacks and Hispanics.

These test-score differences represent real disparities in academic knowledge among groups. Using standards set in the NAEP "main" tests (a set of tests that are updated periodically to allow for change in pedagogy), the latest average scores for twelfth-grade whites in all three subject areas fall between "basic" and "proficient" (Table 1).<sup>18</sup> The average 1998 reading scores for blacks and Hispanics are in this same band, but the 2000 mathematics and science scores fall short of the "basic"

<sup>&</sup>lt;sup>15</sup> As with the data on educational attainment, these numbers come from the Current Population Survey. Results based on the latest decennial Census (2000) are somewhat different because of its much greater coverage of the population, but they also show sharp changes in the composition of the population by race and ethnicity.

<sup>&</sup>lt;sup>16</sup> The share for American Indians and Aleut Eskimos is about 1 percent; this group continues to have low average educational attainment.

<sup>&</sup>lt;sup>17</sup> For the NAEP scores, the white, black, and Hispanic categories are mutually exclusive.

<sup>&</sup>lt;sup>18</sup> More recent NAEP "main" tests have moved towards a greater degree of open-ended questions versus multiple choice and allow greater use of calculators for math problems.



Source: U.S. Department of Education (2000).

#### Table 1

Average Twelfth-Grade NAEP Scores by Race, Hispanic Origin, and Parental Education, versus Standards

	Reading (1998)	Mathematics (2000)	Science (2000)
Race/Ethnicity:			
Whites	298	308	154
Blacks	270	274	123
Hispanics	275	283	128
Parental Education:			
Graduated from college	301	313	157
Some education after high school	292	300	146
Graduated from high school	280	288	135
Did not finish high school	268	278	126
Standards:			
Advanced	346	367	204
Proficient	302	336	170
Basic	265	288	138
Source: U.S. Department of Education (1999c	, 2001b, and 2003).		

achievement level.<sup>19</sup> Indeed, on the four administrations of the mathematics tests given between 1990 and 2000, average black and Hispanic twelfth-grade scores almost always fell short of the basic achievement score of 288.<sup>20</sup> According to the NAEP, this meant that these students did not have a high probability of being able to determine the cost of renting a car given the per-day and mileage charges, nor were they able to apply the concept of perimeter.

To some extent, the test score differences for white, black, and Hispanic high school students reflect differences in family circumstances, such as the disparate educational attainment of their parents' generation. Higher (lower) parental education is associated with higher (lower) student test scores on the NAEP (Table 1). In the 1998 reading test and the 2000 mathematics and science tests, students whose parents had received some education beyond high school had average scores above the "basic" achievement standard. Students whose parents' highest education was a high school degree had average math and science scores that either just barely met the "basic" standard or fell short of it. In light of the differences in schooling completion rates among racial and ethnic groups over the past three decades, a higher percentage of black and Hispanic students are likely to have less-educated parents, which contributes to continuing gaps in NAEP test scores.

Achievement differences in NAEP scores resulting from family background call into question the equalizing effect of public schools. Nevertheless, data on school resources, specifically computer and Internet access, indicate that schools tend to equalize access to information technology compared to what white, black, and Hispanic students have at home.<sup>21</sup> In 2000, Hispanic children aged 6 to 17 were only one-half as likely as whites to have access to a computer at home—38 percent, versus 79 percent (Figure 8). However, 70 percent of Hispanic children reported being able to use a computer at school, compared to 84 percent for whites (Figure 9).

While smaller than the disparities in homes, the school disparities have remained fairly persistent over time. The 14-point gap between white and Hispanic-origin students in computer access at school is similar to the 13-point gap in school use as of 1984. While still below that of white students, black school-age children's recent rates of computer

<sup>&</sup>lt;sup>19</sup> Admittedly, a National Research Council committee concluded that the NAEP cutoffs for "proficient" have been set too high, but they did not draw a similar conclusion with respect to the definition of the "basic" standard (National Research Council 1999).

 $<sup>^{\</sup>rm 20}$  The one exception was Hispanics in 1996, whose scores were just higher than this standard.

<sup>&</sup>lt;sup>21</sup> Supplements to the Current Population Survey in October 1984, 1989, 1993, and 1997 provide information about school-aged children's use of computers, and the supplement in August 2000 provides an update on access (but not actual use). The 2000 data on blacks and whites define these categories exclusive of Hispanics.



Note: Data for 2000 for "whites" and "blacks" do not include Hispanics; in all other years, "whites" and "blacks" include Hispanics.

Source: U.S. Census Bureau (October 1984, 1989, 1993, 1997; August 2000).

usage and access have been slightly above the rates for Hispanics. The 1997 survey shows noticeable convergence between black and white computer usage, but the 2000 survey suggests renewed divergence.

The ongoing computer-access gap in schools among whites, blacks, and Hispanics seems to contradict the widespread publicity over the major strides made in hooking schools up to the Internet since the 1996 "E-rate" legislation.<sup>22</sup> Indeed, 96 percent of public schools with 50 percent or more minority enrollment had Internet access in 2000, more than a 30 percentage point increase since 1997 and not much below the 100 percent rate for schools with very few minority students (Figure 10).<sup>23</sup> However, the percentage of instructional rooms with Internet access has continued to differ sharply across schools with different racial compositions. In 2000, schools with the largest minority enrollments had only 64 percent of instructional rooms wired to the Internet, while schools with little minority representation had 85 percent of rooms hooked up (Figure 11).

The gaps in school and home resources as indicated by technology access and manifested in student test-score data have a lasting effect on the relative achievement levels of whites, blacks, and Hispanics. Post high

 $<sup>^{\</sup>rm 22}$  Officially known as the Universal Service Order provision of the 1996 Telecommunications Act.

<sup>&</sup>lt;sup>23</sup> Minorities include all groups except non-Hispanic whites.



Figure 9 School Computer Use among Children Aged 6 to 17,

Note: Data for 2000 for "whites" and "blacks" do not include Hispanics; in all other years, "whites" and "blacks" include Hispanics. Source: U.S. Census Bureau (October 1984, 1989, 1993, 1997; August 2000).

school education and training do not counteract these effects, as racial achievement differences exist even among persons with similar years of schooling. In 1992, the U.S. Department of Education administered the National Adult Literacy Survey to some 13,500 persons 16 years of age and older.<sup>24</sup> The survey tested respondents on reading comprehension, the ability to use documents such as tables and forms, and the ability to use printed materials to perform computations or other quantitative analysis. The resulting scores were translated into five levels of literacy. No dividing line has been established between literate and illiterate, but income and employment are strongly and positively correlated with literacy scores. Moreover, individuals who demonstrate only level-one or level-two literacy are much more likely to be receiving food stamps and living below the poverty line.25

Not surprisingly, average proficiency increased with years of education. However, within each educational attainment category, literacy scores also varied among whites, blacks, and Hispanics, reflecting

<sup>&</sup>lt;sup>24</sup> This test was re-administered in 2002 with a greater attempt to link demographic and background information with literacy levels. Results are not yet available.

<sup>&</sup>lt;sup>25</sup> Over 40 percent of adults scoring in levels one and two live in poverty, compared to 4 to 8 percent of adults scoring in the highest two levels. Further, 17 to 19 percent of adults in levels one and two receive food stamps, compared to only 4 percent for individuals in levels four and five.



Figure 10 Internet Access in Public Schools, by Percent Minority Enrollment



a persistency in the achievement patterns seen among teenagers (Figure 12).<sup>26</sup> Among "terminal" high school graduates, the average white literacy score was considerably higher than that of blacks, putting the average white adult at literacy level three, while blacks and Hispanics remained at level two. The score gap remained similar between blacks and whites with some college, and increased among college graduates. The average white college graduate was at level-four literacy, while among blacks and Hispanics, only college graduates demonstrated average literacy above the lowest two categories. According to the U.S. Department of Education, these literacy scores imply that the average black and Hispanic adult with less than four years of college has a low probability of being able to use a bus schedule for a given set of conditions and is not likely to be able to interpret instructions for an appliance warranty (1993).

<sup>&</sup>lt;sup>26</sup> For the National Adult Literacy Survey, "white" and "black" include persons of Hispanic origin. Some of the reported gaps for people not currently in high school may include the effects of poorer education for older cohorts of minorities, since they are not broken down by age.





## Educational Attainment by Sex: Persistent Differences in Subject Area Test Scores

In contrast to the patterns among whites, blacks, and Hispanics, male–female differences in educational attainment have largely disappeared over time. At least since 1970, adult women and men have demonstrated equal high school completion rates (Figure 13). Although college completion rates differed greatly for men and women in the 1970s, over the past 30 years this gap has shrunk as a result of steeper increases in college graduation rates among women. By 2000, in the population aged 25 years and older, almost 28 percent of males had completed four years of college compared to 24 percent of women.

The remaining gap in college completion is due to differences between males and females within the white population. Black women's college completion rates traditionally have been on par with black men's, and Hispanic women had matched Hispanic male completion rates by the late 1990s.

Focusing on the younger generation shows that since the mid-1990s, a slightly higher percentage of all women than men in the age group 25 to 29 have completed four years of college (Figure 14). Thus, the gaps remaining in the adult population likely will evaporate over time—or even reverse—as younger women continue to match or exceed younger men in educational attainment.



Source: U.S. Department of Education (1999a).

Figure 13 High School and College Completion Rates, U.S. Population Aged 25 and Over, by Sex



Source: U.S. Census Bureau (March, various years).



Figure 14 High School and College Completion Rates, U.S. Population Aged 25 to 29, by Sex

Although men and women are becoming equal in their likelihood of completing college, differences still exist in assessment test scores by different subject areas. On the NAEP, 17-year-old boys performed better than girls on the math and science tests, while girls outperformed boys on the reading test (Figure 15). The gender gap is no longer considered statistically significant in mathematics, but a meaningful difference continues to exist in science. The gender gap has widened in reading as male scores worsened throughout the 1990s. These persistent achievement differences suggest that men and women may continue to choose different mixes of occupations in future years. These choices may imply different incomes, if mathematical and scientific skills are compensated more or less highly than language skills.

### Occupations: Limited Opportunities for High School Dropouts, Growing Opportunities for College Graduates

Educational attainment is strongly linked to employment opportunities. The past and present inequality in educational attainment and achievement between demographic groups has contributed to differences in the occupational mix of these groups. This, in turn, creates income and employment gaps by race, ethnicity, and gender. Studying changes in the educational attainment levels of workers in major occupation groups suggests that opportunities for high school dropouts are disappearing


throughout the economy. While high school graduates still comprise the majority of workers in most major occupational categories, they do not dominate the fastest-growing occupations. Instead, college graduates are increasingly dominating the fastest-growing fields.

The educational attainment levels of workers in every major occupational category have increased over the past 30 years. In 1970, only a minority of service workers, machine operators, assemblers, inspectors, farm workers, and laborers had completed high school. Now, high school completion is the norm across the board. In each of the nine major occupation groups, 70 percent or more of the workforce has at least a high school diploma (Figure 16). This implies that a high school diploma is a common requirement for most types of jobs.

The percentage of college degrees has also increased within occupational groups, but typical increases have been modest (Figure 17).<sup>27</sup> As recently as 2000, only professional and technical occupations employed a majority of workers with four years of college. The next closest categories were executive, administrative, and managerial (49 percent) and sales (32 percent). Each of the other categories had 15 percent or fewer workers with four years of college. Thus, the workforces across different occupations have remained very different in the prevalence of a college degree.

However, as is well known, the fastest-growing occupations have been the ones that employ college-educated workers most intensively (Figure 18). Professional and technical occupations employed 16.7 percent of the workforce just after a major classification break in 1983 and 19.8 percent in 2000. Executive, administrative, and managerial occupations increased from 11.8 percent to 16.1 percent over this same period. The largest decline was in machine operators and related professions, which have traditionally had a low representation of educated workers whether measured by either high school or college completion. The expansion of occupational fields that employ larger shares of college graduates indicates the growing importance of these degrees. This evidence coupled with the growing necessity of a high school degree illustrates the potential future limits on occupational opportunities for groups who lag behind in educational attainment.

# EDUCATIONAL ATTAINMENT AND EARNINGS EQUALITY

The last section detailed the lingering inequalities in educational attainment in the United States, especially among whites, blacks, and persons of Hispanic origin, and to a much more limited degree between men and women. This section explores the quantitative impacts of these differences on earnings inequality. To what extent do blacks' and His-

<sup>&</sup>lt;sup>27</sup> The classification change in 1983 appears to have been quite significant for changing the percentage with a college degree in some occupations. In particular, executive, administrative, and managerial occupations had a greater presence of college graduates after the classification change than before. As seen later, the sales category grew considerably as a result of reclassification, but its share with a college degree decreased only modestly.







EDUCATIONAL ATTAINMENT AS A CONSTRAINT ON ECONOMIC GROWTH 63



Figure 19 Median Weekly Earnings of Full-Time Workers

panics' lower educational attainment account for their lower economic status? To what extent can male-female income differences be explained by educational differences?

## The Rising Penalties to a Lack of Education

As has been widely acknowledged and analyzed, educational attainment has been of growing importance in determining income, particularly in the United States, which has relatively little regulation or centralized coordination of pay scales compared to most other nations. Less-educated persons tend to be out of work more frequently than highly educated persons. Moreover, during the past couple of decades, even full-time employment has been associated with declining real earnings over time for the less educated. Meanwhile, college graduates have enjoyed a growing payoff to their education.

Figure 19 illustrates the growing earnings differences associated with different levels of educational completion among full-time workers. Among those with less than a four-year college degree, median real earnings fell almost continually from 1979 to the mid-1990s. Adjusted for inflation, median earnings dropped 27.4 percent for those with less than a high school education, 11.7 percent for those with only high school, and 8.3 percent for those with some college. The increases in the late 1990s still

leave these workers' median weekly earnings in 2000 below what they were a decade earlier.

By contrast, pay generally has increased over time for those with a college degree or more, albeit at different rates in different time periods. As a result, in 2000, the median full-time worker with a four-year college education earned 67 percent more than one with only a high school diploma. In 1980, this differential had been 36 percent, or roughly one-half of its current spread.

# Differential Payoffs to Education: An Important Source of Earnings Inequality

The sharp earnings penalty for a lack of education, combined with the growing payoff to completing college, suggests that the lingering differences in college completion rates between whites and blacks, and the growing differences between whites and Hispanics, could have major ramifications for economic inequality. This section will attempt to quantify this effect. To do so, this paper relies heavily on the insights and numerical findings of a recent study by Bradbury (2002).

Bradbury's contribution is to point out that the typical payoffs to further education have varied among demographic groups in the United States. Based on regression analysis using the Current Population Survey, she finds that blacks and Hispanics did not see as steep an increase in the educational wage premium between 1980 and 2000 as their nonblack or non-Hispanic counterparts. Thus, minorities' earnings were held back, not just because they had lower educational attainment levels, but because the payoff to education was not as great as for majority earners.

Although Bradbury focused on changes in the educational wage premium over time, her data and estimated coefficients can be used to answer the following questions: How much of the earnings gap between blacks and nonblacks would be closed if blacks completed the various levels of schooling at the same rates as nonblacks? How much do differences in educational attainment account for earnings differences between Hispanics and non-Hispanics and between women and men? As detailed below, it turns out that "simply" equalizing years of

As detailed below, it turns out that "simply" equalizing years of schooling would close only one-fifth to one-third of the observed earnings gaps between minority and majority men who work full time, and roughly one-half of the earnings gap between minority and majority women who work full time. The remaining earnings gaps result from non-Hispanic whites earning much more for any given level of education than Hispanics or blacks. This suggests that there are earnings penalties associated with a lower-quality education and the other characteristics of minority neighborhoods, or that labor markets discriminate by race and ethnicity, or that some combination of these various factors leads to earnings gaps across groups.

1979–1980	1999–2000
148.81	138.13
156.20	148.91
119.61	122.82
41.68	34.11
147.59	237.72
236.33	282.73
172.77	181.78
64.49	121.51
	1979–1980 148.81 156.20 119.61 41.68 147.59 236.33 172.77 64.49

Table 2 Sources of Weekly Earnings Differences for Men Constant 2000 Dollars

Source: Author's estimates and Bradbury (2002) using "upper-bound" coefficient estimates that exclude occupation and industry from the equations.

Table 2 summarizes the evidence for men, by different racial and ethnic groupings. In 1979–80, the average real weekly wage was \$478 for black full-time male workers and \$627 for nonblack full-time male workers, for a difference of \$149. (These earnings are expressed in 2000 dollars.) Part of this wage difference is associated with factors not directly linked to education, such as usual work hours per week, marital and family status, potential years of work experience, and region of the country. According to Bradbury's regressions, equalizing these other factors produces a slightly larger wage difference, \$156.<sup>28</sup>

To determine the share of the wage difference caused by educational

<sup>&</sup>lt;sup>28</sup> The simulations reported here use Bradbury's "upper-bound" estimates for education. Occupation and industry mix are omitted from the independent variables in the regressions. Thus, whatever added differences in earnings may be attributable to occupation and industry are subsumed in the other coefficients, and the simulations do not explicitly equalize the mix of occupations and industries. Bradbury's "lower-bound" estimates include occupation and industry as separate regressors. Using these results and equalizing occupation and industry choices across groups changes the numerical conclusions somewhat for minority versus majority women, but hardly at all for men. Nor does this assumption matter in analyzing overall male–female differences. Appendix B presents the full details of the two sets of estimates.

differences, a new set of calculations was performed in which black men were assigned the same educational attainment patterns as nonblack men. The percentage of black men completing less than high school dropped dramatically in this simulation, while the percentages completing only high school, some college, college, and more than college were each increased. The new educational attainment rates were multiplied by the estimated payoffs for black men from completing each level of education, as estimated in Bradbury's regressions. The increase in black men's years of schooling lifted their weekly earnings by about 8 percent, or \$37. However, this increase was only 23 percent of the simulated earnings gap in 1979–80. Similarly, when the same exercise was done using the 1999–00 observations on educational attainment and payoffs to education, black men's weekly earnings rose about 6 percent. The weekly earnings gap narrowed by \$26, only 18 percent of the \$149 simulated earnings gap between black and nonblack men.

In a second new simulation exercise, black men retained their actual composition of educational attainment, but each educational attainment level was assumed to earn the same return in the labor market as that experienced by white men. In her paper, Bradbury found that, holding constant a range of other attributes, nonblack high school graduates currently earn about 20 percent more than black high school graduates, while nonblack college graduates earn 23 percent more than black college graduates. She found similar differences in 1979–80.<sup>29</sup> Not surprisingly then, giving black men the nonblack earnings at each level of education would raise their simulated earnings considerably. Indeed, black men's real weekly earnings were raised by about \$115 in both 1979–80 and 1999–00. This amounted to roughly three-quarters of the observed earnings gap between the two groups, holding non-education factors constant.

Performing the simulation exercises for Hispanic versus non-Hispanic males yields broadly similar results. Because Hispanics have fallen further behind non-Hispanics in their average years of schooling, raising their years of schooling closes more of the earnings gap than is the case for blacks. However, the shortfall in quantity of schooling still does not account for the bulk of their shortfall in earnings. If Hispanic men had achieved the non-Hispanic men's educational mix, the size of their simulated real earnings gaps would have been reduced from \$236 to \$173, or by 27 percent, in 1979–80. In 1999–00, 36 percent of their earnings gap would have been closed. As in the case of black men, a far larger share of

<sup>&</sup>lt;sup>29</sup> See Bradbury's Figure 12 for differences in payoffs for blacks and nonblacks and Figure 13 for Hispanics and non-Hispanics. The numbers cited rely on the "lower-bound" estimates, but the "upper-bound" estimates are not very different.

Constant 2000 Dollars					
	1979–1980	1999–2000			
Black Women versus Nonblack Women					
Actual Difference Simulated difference if blacks given	27.93	62.94			
nonblacks' characteristics and: If each group's education mix and returns to					
education kept at actual values	34.93	58.07			
If blacks given nonblack education mix If blacks given nonblack returns to	19.56	33.77			
education	14.70	25.55			
Hispanic Women versus Non-Hispanic Women Actual Difference	53.82	131.45			
Simulated difference if Hispanics given non- Hispanic characteristics and:					
education kept at actual values	73.31	131.77			
mix If Hispanics given non-Hispanic returns to	32.46	54.28			
education	33.01	80.26			
All Women versus All Men					
Actual Difference	207.53	141.03			
Simulated difference if women given men's					
characteristics and:					
If each group's education mix and returns to					
education kept at actual values	159.45	86.13			
It women given men's education mix	159.56	98.70			
IT women given men's returns to education	-3.26	-13.29			

#### Table 3 Sources of Weekly Earnings Differences for Women Constant 2000 Dollars

Source: Author's estimates and Bradbury (2002) using "upper-bound" coefficient estimates that exclude occupation and industry from the equations.

the earnings gap is accounted for by lower labor market returns from completing given amounts of education.

Turning to women who work full time, the earnings differences between blacks and nonblacks and between Hispanics and non-Hispanics are much smaller—on the order of one-half of those among men in 1999–00 (Table 3). Additionally, the returns to education are more similar among women of different racial and ethnic groups than among men. For example, black (Hispanic) female high school graduates working full time earn roughly one-tenth less than nonblacks (non-Hispanics); the percentage gaps are similar among college graduates.<sup>30</sup> As a result, providing black women with the same number of years of schooling as nonblack women closes their current earnings gap by 42 percent, while providing them the same rates of return for a given number of years of education closes the gap by about 56 percent. For Hispanics versus non-Hispanics, the percentages are roughly the reverse: Raising the number of years of schooling for Hispanics would reduce their earnings gap by more than half.<sup>31</sup>

Finally, a similar exercise was conducted for comparing women and men. These two groups' years of schooling are quite similar. As discussed earlier, women lag behind men somewhat in four-year college completion. However, among full-time workers, greater shares of women than men complete high school and get some education beyond high school. Not surprisingly then, all of the earnings difference between men and women who work full-time can be attributed to differences in the returns from completing a given level of education (after equalizing weekly hours of work, the influences of family and marital status, and the other variables used to produce simulated earnings).

## **Implications for Policy**

Trying to attribute the observed earnings gaps among whites, blacks, and Hispanics into separate portions categorized by differences in years of schooling and differences in returns to completing a given number of years of schooling is somewhat artificial. If blacks and Hispanics were able to earn the same amount upon completion of high school or college as whites, they would likely stay in school longer.<sup>32</sup> Moreover, the analysis in Bradbury's study and this paper considered only full-time workers. Persons who receive little education are less likely to be in the workforce and less likely to be employed, relative to their more-educated peers. Thus, raising the years of schooling for blacks and Hispanics would tend to have an additional equalizing effect on earnings by raising their likelihood of being employed, which is not measured here.

Nevertheless, the analysis strongly suggests that, to combat the earnings gap, more emphasis should be put on policies that raise the payoff to education for minority groups. The first step in this process is developing an understanding of why returns from completing the same number of years of schooling differ across population groups. After

<sup>&</sup>lt;sup>30</sup> These percentages are from Bradbury's "lower-bound" estimates shown in her Figures 12 and 13; the simulations reported here use the "upper-bound" estimates.

<sup>&</sup>lt;sup>31</sup> Hispanic women's educational attainment levels are further below non-Hispanic women's than is the case for black women relative to nonblack women.

<sup>&</sup>lt;sup>32</sup> See Cameron and Heckman (2001) for further discussion of incentive effects and barriers to educational attainment for blacks and Hispanics.

reviewing a variety of empirical studies, Bradbury concludes that the differentials reflect a combination of influences. First, institutional factors, such as labor market discrimination, create distinctions between groups. Second, "differences exist in the quality of education obtained by different groups, implying that individuals with similar 'educational attainment' do not actually have the same education, and, by implication, job skills" (p. 41). Indeed, this paper has shown evidence of differences in school resources, standardized test scores, and literacy among racial and ethnic groups. An additional problem is likely to be geographic segregation by income and race, which leaves many minority households living in neighborhoods without established job networks and far away from fast-growing suburban employers (Bradbury, Kodrzycki, and Mayer 1996). Since white women tend to live in the same locations and attend the same schools as white men, the earnings differences associated with similar educational attainment are likely reflective of differences in career paths.

# SHORTFALLS OF TECHNICAL TALENT AND THEIR IMPLICATIONS

A final set of concerns for the United States is that the skill mix of the educated labor force may be suboptimal in some sense. If the mix of knowledge embodied in workers is out of line with the demands of employers, economic growth may be curtailed in the affected industries, perhaps enough to spill over to the economy as a whole.

Although these concerns have waxed and waned over time, they keep reemerging and usually focus on scientific and technological skills. For example, as early as the 1950s, studies appeared on engineering shortages in the United States, and in the late 1980s, projections for a shortfall of engineers during the 1990s became commonplace. More recently, in the second half of the 1990s, employers perceived a shortage of information technology workers, not only in the United States but worldwide, prompting the National Research Council to commission a detailed, high-level study of these issues (National Research Council 2001).

The focus of concerns on technical occupations is, in part, a consequence of their perceived importance in overall economic growth and in achieving additional national objectives. For example, a study issued by the U.S. Bureau of Labor Statistics (Braddock 1992) began as follows:

Our Nation's economic progress and general well-being depend in considerable measure on the work of scientists, engineers, and technicians. These men and women contribute to the development of new products, improvements in productivity, enhanced defense capabilities, environmental protection, and advances in communications and health care (p. 28). If anything, this perception has increased as technology's role in recent economic growth has been emphasized:

It is important to the nation that there be an adequate number of scientists and engineers. Industries that rely on scientific and technical research and development are increasingly important in both the global and American economies. If there are too few scientists and engineers, the economy and its competitive position, both now and in the future, are put at risk (National Research Council 2000, p. 15).

# Supply or Demand?

From the standpoint of economic theory, a shortage can develop in the short run as the relative demand for different skills shifts and the supply of appropriately skilled workers does not match the demand. In response, wages or other aspects of compensation for these skills increase so that shortages tend to be eliminated over time.<sup>33</sup> However, the mismatch of skills may pose a longer-term problem if demand spikes unexpectedly for skills that are acquired only over a lengthy period of education or training, or if market barriers prevent wages from adjusting.<sup>34</sup> It is important to examine the mechanics of technical labor markets to assess the potential danger of longer-term shortages.

Little if any evidence exists that shortages of scientific and technical workers are a permanent feature of advanced economies such as the United States. For this to be true, the private return (that is, wages and other forms of compensation) in these occupations would have to fall short of the productive contributions of the workers on a continual basis. Despite the acknowledgement of the importance of scientific and technical innovations, no study has yet indicated that market failures cause these workers to be underpaid.<sup>35</sup> Developing countries, by contrast, may face a chronic "brain drain" problem as skilled professional and technical workers migrate to more advanced countries, where pay tends to be higher.

Instead, tightness in scientific and technical fields tends to be episodic. Demand for these skills, on occasion, has shifted abruptly and for an unpredictable period of time as a result of policy or technological

<sup>&</sup>lt;sup>33</sup> The other aspects of compensation may include monetary benefits or nonmonetary amenities such as improved working conditions or enhanced prestige.

<sup>&</sup>lt;sup>34</sup> Another possible problem might be if demand for certain skills somehow chronically increases too much for supply to adjust. This possibility has been modeled theoretically, but it has not received empirical support.

<sup>&</sup>lt;sup>35</sup> Within the U.S. context, it may plausibly be argued that teachers currently are underpaid, since this field has been dominated by women, whose professional opportunities have been limited historically as a result of sex discrimination and social norms. See Temin (2002).

change. For example, the demand for engineers and other technical workers rose considerably from the late 1970s to the late 1980s as U.S. defense procurement outlays doubled as a share of GDP. Recently, demand for information technology (IT) workers rose sharply as real investment in information processing equipment and software went from 3 percent of real GDP in early 1995 to almost 7 percent by the end of 2000.<sup>36</sup> The college labor market as a whole is not faced with such sharp swings in demand. As noted by Ryoo and Rosen (2001), most other types of college-educated workers tend to be employed in relatively stable services industries.<sup>37</sup>

These sudden demand shifts combine with slow supply adjustment to create potential problems in technical fields. Engineers and computer scientists must receive appropriate education and training. This requires not only a significant amount of time, but also flexibility within educational institutions to adjust their instructional staff and facilities. However, it is not clear that this slow supply adjustment is unique to technical fields. At a first approximation, the adjustment periods for these fields are likely similar to those for other occupations that are dominated by a highly educated workforce.<sup>38</sup>

It is plausible that the historic volatility of demand in technical fields may lead prospective workers to discount wage and salary signals, slowing the supply adjustment relative to workers in other steadier fields. Following the defense buildup of the 1980s, demand for engineers and technical workers was halted by the end of the Cold War and the resulting dramatic declines in defense procurement. The boom of spending on IT in the late 1990s was followed by a dramatic bust that brought on a national recession and resulted in lower demand for IT workers.<sup>39</sup>

However, contrary to ex post evidence concerning demand volatility in technical occupations, the National Research Council's report, issued

<sup>&</sup>lt;sup>36</sup> Admittedly, these statistics on expenditures are not indicative of demand alone, but also reflect the supply of workers in the industries producing these goods and services.

<sup>&</sup>lt;sup>37</sup> Moreover, because the upswings in demand for technical workers have been so strong on occasion, they have contributed significantly to national economic growth. Thus high demand for engineers in the late 1980s and high demand for IT workers in the late 1990s coincided with periods of low overall unemployment, which compounded the recruitment and retention of these workers.

<sup>&</sup>lt;sup>38</sup> Moreover, many positions in technical fields can be filled with persons with a limited period of formal education. The National Research Council estimated that about one-half of the five million positions in information technology fields involve the application, adaptation, configuration, support, or implementation of IT products designed or developed by others. These positions do not require lengthy formal education and training periods. Of the higher-level positions involving development of IT products, about two-thirds of the workers have at least a bachelor's degree, but in many cases their university degrees were in fields other than computer science, which suggests that graduates in other fields were able to retrain for IT (National Research Council 2001).

<sup>&</sup>lt;sup>39</sup> For example, Internet advertisements for high-technology workers in New England fell 75 percent between early 2001 and early 2002 (*Mass High Tech* 2002).



Figure 20 **Total Bachelor's Degrees Conferred and Shares in Engineering** 

Source: U.S. Department of Education (2002).

before the downturn was evident, surmised that "there is some historical precedent for thinking that the IT sector might be affected less severely than other sectors by an overall downturn and even that IT growth can continue during an overall downturn" (2001, p. 119). Thus, volatility could have dissuaded students from pursuing IT-related degrees only if they were more farsighted than objective experts were.

Moreover, a look at recent trends in college majors suggests that escalating salaries for IT specialists have elicited a supply response. The share of U.S. bachelor's degrees awarded in computer and information sciences rose from about 2 percent in the mid-1990s to about 3 percent in 2000 (Figure 20). Conversely, engineering's share of bachelor's degrees has fallen continuously since the late 1980s, despite Ryoo and Rosen's estimate that "the speed of response in this market to changing conditions is rapid" (p. 2).

One problem, identified by Romer (2000), may be that engineering schools do not advertise engineering salaries to their prospective students to the same extent that business and law schools do. If prospective students do not realize that the relative pay for engineers has risen, this would tend to lengthen the adjustment period following an increase in demand for engineers. By contrast, the abundance of Internet-based salary information for IT positions may lead to a relatively faster

Figure 21 Share of Bachelor's Degrees Conferred to Women in All Subjects, Computer Sciences, and Engineering



Source: U.S. Department of Education (2002).

adjustment to changes in pay.<sup>40</sup> In addition, however, the rates of increase in IT salaries, at least in the late 1990s, appear to have been higher than for engineers (National Research Council 2001).

Another barrier to increasing the share of college students completing engineering degrees may be the continuing under-representation of young women. Even as the share of all U.S. bachelor's degrees awarded to women approaches 60 percent, the share of engineering bachelor's degrees awarded to women remains under 20 percent (Figure 21). Although computer science also has a preponderance of male majors, its female share has been considerably higher than that in engineering. The continuing weaker performance of high school girls than boys on science tests may exacerbate the challenge of shifting a greater share of college students into technical fields.

## Demographics as a Current Complication for Supply

This study has argued that, given the mixed evidence on supply adjustment, it is the sudden demand shifts relative to other sectors that

<sup>&</sup>lt;sup>40</sup> See National Research Council (2001) for examples of web sites.



Source: U.S. Department of Education, (2002).

are especially important in creating periodic tightness or shortages in scientific and technical fields. However, the upsurge in demand for IT workers in the 1990s took place against a backdrop of constraints on supply that appear to be both predictable and longer-lasting.

One constraint has been the slowing increase in college attendance among the young. Another constraint has been slow population growth among the age group that typically attends college. Because of the maturing of the baby boom, the share of 18- to 24-year-olds in the total U.S. population 18 years and over fell from about 19 percent in the late 1970s to about 13 percent in the late 1990s (Figure 22).

As discussed above, demographically based projections call for only modest increases in the number of college graduates during the next two decades (Ellwood 2001). To the extent that projections of a reemergence of growth in demand for IT workers come to pass, the anticipated overall slow growth in supply of college-educated workers would tend to constrain the ability to fill positions—*even if* choices of college majors are responsive to market signals. Thus, mechanisms to retrain the adult workforce as demand for technical skills increases appear to deserve even greater attention than in the past.

# SUMMARY AND CONCLUSIONS

This study provides support for the view that the existing patterns of educational attainment in the United States threaten social progress. The Hispanic share and, to some extent, the black share of the population have risen over time. Yet not only do blacks and Hispanics complete fewer years of schooling than whites, but more important for their economic status, their educational resources and achievement lag behind at each level of schooling. Some of these gaps, particularly performance on standardized tests, have widened in the past decade. Moreover, newer educational initiatives, such as providing access to information technology in the classroom, have been introduced less widely in schools with higher proportions of minority students. These apparent differences in the quality of schooling received by whites, blacks, and Hispanics, as well as likely differences in non-school inputs and access to jobs, account for a greater share of earnings differences observed among full-time workers in these groups than differences in their years of schooling.

Because children's educational achievement has been closely linked to the levels of education completed by their parents, raising educational achievement for racial and ethnic minorities will take a sustained effort. By contrast with the differences by race and ethnicity, differences in the economic status between men and women are no longer attributable to differential rates of access to higher education. Instead, they are likely associated with continuing differences in the fields of work that men and women pursue.

Unless new policies offset the effects of existing demographic and educational patterns, improvements in labor quality are likely to contribute less to economic growth in the United States in the coming two decades than has been the case since the 1960s. The key reasons for this projection are the relatively slow increase in years of schooling obtained by young adults and the relatively low share of the population in the age group when labor market entry typically occurs. Furthermore, international test scores indicate a continuing mediocre performance for U.S. students on average. These trends suggest that capital formation or technology development would have to provide an offset in order to keep per capita income growth from slowing in coming decades. They also suggest that surges in demand for educated labor, as have occurred periodically in scientific and technical fields, will be challenging to accommodate.

#### References

Acemoglu, Daron. 1998. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *Quarterly Journal of Economics* 113 (4): 1055–90.
Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal*

Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." Quarterly Journal of Economics 106 (2): 407–43. ——. 2001. "Human Capital and Growth." American Economic Review 91 (2): 12–17.

Barro, Robert J. and Jong-Wha Lee. 2000. "International Data on Educational Attainment Updates and Implications." NBER Working Paper No. 7911 (September).

Barro, Robert J. and Xavier Sala-i-Martin. 1995. Economic Growth. New York: McGraw-Hill.

- Benhabib, Jess and Mark M. Spiegel. 1994. "The Role of Human Capital in Economic Development: Evidence From Aggregate Cross-Country Data." *Journal of Monetary Economics* 34 (2): 143–73.
- Bernanke, Ben S. and Refet S. Gürkaynak. 2001. "Is Growth Exogenous? Taking Mankiw, Romer, and Weil Seriously." NBER Working Paper No. 8365 (July).
- Bils, Mark and Peter J. Klenow. 2000. "Does Schooling Cause Growth?" American Economic Review 90 (5): 1160–83.
- Boesel, David, Nabeel Alsalam, and Thomas M. Smith. 1998. "Educational and Labor Market Performance of GED Recipients." <a href="http://www.ed.gov/pubs/GED/title.html">http://www.ed.gov/pubs/GED/title.html</a> 24 May 2002.
- Bradbury, Katharine L. 2002. "Education and Wages in the 1980s and 1990s: Are All Groups Moving Up Together?" New England Economic Review Q1: 19-46.
- Bradbury, Katharine L., Yolanda K. Kodrzycki, and Christopher J. Mayer. 1996. "Spatial and Labor Market Contributions to Earnings Inequality: An Overview." *New England Economic Review* May/June: 1–10.
- Braddock, Douglas J. 1992. "Scientific and Technical Employment, 1990–2005." Monthly Labor Review 115 (2): 28–41.
- Cameron, Stephen V. and James J. Heckman. 1993. "The Nonequivalence of High School Equivalents." *Journal of Labor Economics* 11 (1): 1–47.
- ———. 2001. "The Dynamics of Educational Attainment for Blacks, Hispanics, and Whites." Journal of Political Economy 109 (3): 455–500.
- Clark, Melissa A. and David A. Jaeger. 2002. "Natives, the Foreign-Born and High School Equivalents: New Evidence on the Returns to the GED." IZA Discussion Paper No. 477 (April).
- de la Fuente, Angel and Rafael Doménech. 2000. "Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?" CEPR Working Paper No. 2466 (May).
- Ellwood, David T. 2001. "The Sputtering Labor Force of the 21<sup>st</sup> Century: Can Social Policy Help?" NBER Working Paper No. 8321 (June).
- Freeman, Richard B. 1976. The Overeducated American. New York: Academic Press.
- Glaeser, Edward L., José A. Scheinkman, and Andrei Shleifer. 1995. "Economic Growth in a Cross-Section of Cities." NBER Working Paper No. 5013 (February).
- Hanushek, Eric A. and Dennis D. Kimko. 2000. "Schooling, Labor-Force Quality, and the Growth of Nations." *American Economic Review* 90 (5): 1184–1208.
- Ho, Mun S. and Dale W. Jorgenson. 1995. "The Quality of the U.S. Work Force, 1948–95." Kennedy School of Government, Harvard University, working paper (tables updated February 1999).
- Jorgenson, Dale W., Frank M. Gollop, and Barbara M. Fraumeni. 1987. *Productivity and U.S. Economic Growth*. Cambridge, MA: Harvard University Press.
- Jorgenson, Dale W., Mun S. Ho, and Kevin J. Stiroh. 2002. "Information Technology, Education, and the Sources of Economic Growth across U.S. Industries." Harvard University, working paper (April).
- Krueger, Alan B. and Mikael Lindahl. 2001. "Education for Growth: Why and For Whom?" Journal of Economic Literature 39 (4): 1101–36.
- Kyriacou, George. 1991. "Level and Growth Effects of Human Capital." C.V. Starr Center, Working Paper No. 91-26, New York, NY.
- Lee, Jong-Wha and Robert J. Barro. 1997. "Schooling Quality in a Cross Section of Countries." NBER Working Paper No. 6198 (September).
- Little, Jane Sneddon and Robert K. Triest. 2002. "The Impact of Demographic Change in U.S. Labor Markets." *New England Economic Review* Q1: 47–68.
- Lucas Jr., Robert E. 1988. "On the Mechanics of Economic Development." Journal of Economics 22 (February): 3-42.
- Mankiw, Gregory N., David Romer, and David N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107 (2): 407–37.
- Mass High Tech. March 2002. "Pulse of Technology: A Quarterly Analysis for N.E.

Decision-Makers in the Innovative Sector." <www.masshightech.com/pulse.html>28 May 2002.

- Murnane, Richard J. and Frank Levy. 1996. *Teaching the New Basic Skills: Principles for Educating Children to Thrive in a Changing Economy*. New York: Martin Kessler Books, The Free Press.
- National Commission on Excellence in Education. 1983. A Nation at Risk: The Imperative for Educational Reform. <www.ed.gov/pubs/NatAtRisk> 13 February 2002.
- National Research Council. 1999. Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress, edited by J. W. Pellegrino, L. R. Jones, and K. J. Mitchell. Washington, DC: National Academy Press.
  - 2000. Forecasting Demand and Supply of Doctoral Scientists and Engineers: Report of a Workshop Methodology. Washington, DC: National Academy Press.
- ——. 2001. Building a Workforce for the Information Economy. Washington, DC: National Academy Press.
- Oliner, Stephen D. and Daniel E. Sichel. 2000. "The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?" Federal Reserve Board Working Paper (May).
- Romer, Paul M. 2000. "Should the Government Subsidize Supply or Demand in the Market for Scientists and Engineers?" NBER Working Paper No. 7723 (June).
- Ryoo, Jaewoo and Sherwin Rosen. 2001. "The Engineering Labor Market." University of Pennsylvania and University of Chicago, working paper (October).
- Sala-i-Martin, Xavier. 1997. "I Just Ran Four Million Regressions." NBER Working Paper No. 6252 (November).
- Simon, Curtis J. 1998. "Human Capital and Metropolitan Employment Growth." Journal of Urban Economics 43: 223–43.
- Temin, Peter. 2002. "Teacher Quality and the Future of America." NBER Working Paper No. 8898 (April).
- U.S. Census Bureau. Current Population Survey 2000; August and March supplements; October supplement 1984, 1989, 1993, and 1997.
- U.S. Department of Education. National Center for Education Statistics. 1993. "Executive Summary," in Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey. NCES 1993-275, by I. S. Kirsch, A. Jungeblut, L. Jenkins, and A. Kolstad. Washington, DC.
- . 1999a. Adult Literacy and Education in America: Four Studies Based on the National Adult Literacy Survey. NCES 1999-469, by C. F. Kaestle, A. Campbell, J. D. Finn, S. T. Johnson, and L. H. Mikulecky. Washington, DC.
- ———. 1999b. Literacy in the Labor Force: Results from the National Adult Literacy Survey. NCES 1999-470, by A. Sum. Washington, DC.
- 1999c. The NAEP 1998 Reading Report Card for the Nation and the States. NCES 1999-500, by P.L. Donahue, K.E. Voelkl, J. R. Campbell, and J. Mazzeo. Washington, DC.
- ——. 1999d. *The NAEP Guide: A Description of the Content and Methods of the 1999 and 2000 Assessments.* NCES 2000-456, edited by N. Horkay. Washington, DC.
- 2000. NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance. NCES 2000-469, by J.R. Campbell, C.M. Hombo, and J. Mazzeo. Washington, DC.
- ——. 2001a. Internet Access in U.S. Public Schools and Classrooms: 1994–2000. NCES 2001–071, by A. Cattagni and E. Farris. Washington, DC.
- ——. 2001b. The Nation's Report Card: Mathematics 2000. NCES 2001-517, by J.S. Braswell, A.D. Lutkus, W.S. Grigg, S.L. Santapau, B. Tay-Lim, and M. Johnson. Washington, DC.
- L. Jerry, and A. Lutkus. Washington, DC.
- ——. 2002. Digest of Education Statistics, 2001. NCES 2002-130, by T. D. Synder and C. M. Hoffman. Washington, DC.
  - —. 2003. The Nation's Report Card: Science 2000. NCES 2003–453, by C. Y. O'Sullivan, M. A. Lauko, W. S. Grigg, J. Qian, and J. Zhang, Washington, DC.

# Appendix A: Education as an Explanation for Why Countries Grow at Different Rates

This Appendix provides a more detailed review of the approaches and findings of the endogenous growth literature.

#### Education as a Precursor to Growth

Barro (1991) tested the endogenous growth model in a study of growth rates in per capita GDP for a sample of 98 countries for the period 1960 to 1985. He explored the relationship of per capita GDP growth to initial levels of per capita GDP and human capital, controlling also for a range of other economic and political variables such as the ratio of government consumption to GDP, the degree of political instability, and economic distortions.<sup>41</sup>

Barro concluded that higher human capital levels (holding initial GDP and other variables fixed) are strongly positively related to subsequent growth. Barro also explored the mechanisms by which higher human capital may lead to higher growth. He found, empirically, that countries with high human capital have low fertility rates and high rates of physical investment, both of which tend to raise per capita income growth. He indicated that the regressions help account for the high rates of economic growth in Pacific Rim countries, which had relatively high levels of human capital compared to initial GDP. However, the model fails to account for much of the relatively weak performance over this period for countries in sub-Saharan Africa and Latin America.

A shortcoming of the study is the crude measurement of human capital. Barro used primary and secondary school enrollment rates—that is, total numbers enrolled in school relative to the population size of the relevant age groups. At best, this measure approximates the rate of investment in human capital; it does not indicate the stock of human capital embodied in the working-age population.<sup>42</sup>

Since the publication of Barro's 1991 study, Barro and Lee (2000) have collaborated on improving the measurement of human capital. The newer data are based on combining periodic census or survey measures of the education levels of the adult population and measures of new school entrants, which affect adult education with the appropriate time lags. Representative results presented in Barro and Sala-i-Martin (1995) indicate that average years of secondary and higher schooling have positive impacts on a country's subsequent growth.

The Barro-type analysis has been extended to U.S. cities and metropolitan areas by Glaeser, Scheinkman, and Shleifer (1995) and by Simon (1998). These authors found that areas with higher initial education tended to show higher rates of growth in per capita income and/or population in subsequent decades. Glaeser, Scheinkman, and Shleifer found the presence of high school graduates to be more important than college graduates, whereas Simon attached greater importance to the college-educated labor pool.

#### Human Capital: Analogies to Physical Capital

The main purpose of Mankiw, Romer, and Weil (1992) was to test the endogenous growth model (as supported empirically by the Barro studies) against the neoclassical growth model. Mankiw, Romer, and Weil amended the traditional neoclassical growth model by adding the stock of human capital as a separate factor of production. Thus, physical capital, human capital, labor (measured essentially as the number of workers), and

<sup>&</sup>lt;sup>41</sup> Barro's 1991 study is explicitly empirical. He does not specify a theoretical model of growth. The macro growth literature tends to include initial per capita income in order to test theories about income convergence across countries; these tests are not the focus of this current review of the literature.

<sup>&</sup>lt;sup>42</sup> Basic problems include possible measurement errors in enrollments as well as the ambiguous definition of the denominator, especially for developing countries where students frequently intersperse periods of school attendance with extended periods of absence from school. Nonetheless, Barro's fundamental results held even when the sample was restricted to the 55 countries that had per capita GDP above \$10,000 in 1960.

an exogenously determined level of technology determine the level of output. Using an assumed Cobb-Douglas production technology and making use of steady-state properties of the neoclassical growth model, Mankiw, Romer, and Weil showed that the major determinants of a country's growth in output per capita are its initial per capita output level and the rates of accumulation of physical and human capital.<sup>43</sup> Using the same 98 countries and 1960–1985 time period as Barro (1991), Mankiw, Romer, and Weil concluded that per capita GDP growth varies positively with investments in both human and physical capital. They proxied human capital accumulation by the ratio of secondary school enrollment to the working-age population. In a follow-up to the Mankiw, Romer, and Weil paper, Bernanke and Gürkaynak (2001) rejected certain key findings of the neoclassical model, but concluded nevertheless that a country's rate of economic growth is correlated with its rate of human capital accumulation.

#### Who Is Correct: Barro, or Mankiw, Romer, and Weil?

Benhabib and Spiegel (1994) set out to test the competing views of how human capital affects growth. Is human capital an "ordinary" input akin to labor and physical capital, as in Mankiw, Romer, and Weil, or does it induce growth by facilitating the development or adoption of technology, as in Barro (1991)?

Benhabib and Spiegel started by positing an aggregate Cobb-Douglas production function with physical capital, human capital, labor, and technology as inputs. They estimated the level of human capital of the labor force as a function of 15-year lags of the enrollment rate in primary schools and five-year lags of the enrollment rates in secondary schools and higher education (derived from Kyriacou 1991). The model was estimated for per capita income growth from 1965 to 1985 using a sample of up to 78 countries. Whether measured by the new Benhabib and Spiegel variable or by the cruder variables used in Barro (1991), human capital was found to be insignificant in determining per capita growth (and, in fact, entered negatively).

In another exercise, Benhabib and Siegel tested whether human capital facilitates technological progress (rather than serving as a separate input into production) through two separate mechanisms. Their regressions included the level of human capital as an indicator of a country's capacity for innovation. To indicate the country's capacity for technological catch-up, they also included an interaction term between the level of human capital and the gap between a country's per capita income and that of the leading nation.<sup>44</sup> The results supported the view that human capital is a determinant of growth through the latter mechanism, technology catch-up. By splitting the sample into three separate groups of countries, Benhabib and Spiegel found that this channel is especially important for countries at low levels of economic development. Krueger and Lindahl (2001), discussed below, also concluded that "the positive effect of the initial level of education on growth seems to be a phenomenon that is confined to low-productivity countries" (p. 1130).

#### Measurement Can Make the Difference

Krueger and Lindahl injected a microeconomics perspective into the debate. They noted that the microeconomic (or "Mincer") model of earnings determination posits that an individual's wage is a positive function of years of schooling. This model has been shown to provide a good description of wage differences across individuals in a variety of studies using data for many different nations.<sup>45</sup> Aggregating over individuals within a country and differencing across years yields a macroeconomic version of the Mincer model in which the

<sup>&</sup>lt;sup>43</sup> More precisely, the model determines output per effective worker, which is a function of the number of workers and the level of technology. In their empirical work, Mankiw, Romer, and Weil use output per person of working age.

<sup>&</sup>lt;sup>44</sup> In the discussion of the structural model presented in their Table 5, Benhabib and Spiegel are not clear as to whether human capital and the income gap are measured at the beginning of the sample period.

<sup>&</sup>lt;sup>45</sup> These studies also control for individual differences in labor market experience, sex, and race.

growth in average earnings depends on the change in average education. If the rate of return increases secularly over time, initial education also will enter positively.<sup>46</sup>

Given the apparent success of the Mincer model, Krueger and Lindahl found it puzzling that influential macroeconomic studies conclude that the change in a country's human capital does not matter in determining income growth. One possible explanation is that the degree of education an individual receives is merely an indicator of the individual's (unobserved) ability, rather than something that adds to his or her productive capacity. However, Krueger and Lindahl cited a series of microeconomic studies rejecting the view that education is principally a signaling device.

An alternative explanation, which Krueger and Lindahl support, is that measurement problems prevent schooling changes from entering significantly. They showed that the schooling data developed in Kyriacou and used in studies such as Barro and Lee are poorly correlated when expressed as changes in educational attainment within countries over intervals of time. Furthermore, these data sets appear especially deficient in measuring the amount of secondary and higher education, when compared to the recent World Values Survey. Krueger and Lindahl questioned the inclusion of physical capital formation in growth regression, preferring a more parsimonious specification.<sup>47</sup> When capital is omitted from the regression, they found that the change in schooling is more likely to be significant. Furthermore, its significance was greater when the time period analyzed is increased from five years to 10 or 20 years, which the authors interpret as further evidence of measurement error. Over short periods of time, variations in average schooling data for a country are likely to reflect measurement problems more than true changes in schooling.

In their study, de la Fuente and Doménech (2000) also developed evidence that measurement error has biased the findings of previous studies. The authors constructed new data on educational attainment in the 21 OECD countries for the period 1960–1990, making use of a greater amount of national information and fixing artificial breaks in the series caused by changes in classification criteria.

Additionally, de la Fuente and Doménech posited an aggregate production function in which the output per employed worker depends on the stock of physical capital and the average number of years of schooling of the adult population. They used pooled data at five-year intervals and estimated the equation in both levels and changes. They allowed for time and country dummies. In the equations for the growth rate of per-worker output, the growth in schooling has a significant positive effect when measured by the revised de la Fuente and Doménech data but not when measured according to Barro and Lee.

#### Allowing for Quality Differences in Education

The literature summarized so far has made use of human capital stock measures based on cumulating historical data on school enrollments. Since years of schooling are likely not to be comparable across countries, some very recent studies have made attempts to measure the quality of education received. These studies may be of particular interest to developed nations that have well-educated workforces as measured conventionally but that are increasingly concerned with improving academic achievement.

Hanushek and Kimko (2000) construct composite measures of labor force quality for 31 countries based on six mathematics and science tests administered between 1965 and 1991 by the International Association for the Evaluation of Educational Achievement (IEA) and the International Assessment of Educational Progress (IAEP). They extended the analysis to a sample of about 80 countries by constructing labor force quality measures via

<sup>&</sup>lt;sup>46</sup> Macroeconomic studies use the change in log GDP per capita as the dependent variable, not the change in the mean of log earnings. Krueger and Lindahl indicate that if income has a lognormal distribution over time, and if labor's share is constant, the results from these two alternative dependent variables should be the same.

<sup>&</sup>lt;sup>47</sup> One issue is endogeneity: Fast growing countries may have greater access to capital. Another issue is artificial correlation with growth, since capital formation is derived from data on investment, which is a component of GDP. Barro (1991) did not include capital formation among the independent variables.

regression analysis using the limited test scores for some countries, along with additional indicators such as family characteristics and school resources.

Hanushek and Kimko performed regression analysis to explain differences in crosscountry growth rates during the period 1960 to 1990. They found that the quantity of schooling (as measured by Barro and Lee) becomes insignificant when the labor force quality measures are added. Furthermore, the inclusion of labor force quality substantially boosts the explanatory power of the regressions.

A closely related panel study by Barro (2001) examined per capita GDP growth for 100 countries in three time periods: 1965–1975, 1975–1985, and 1985–1995. Quality is measured by science, mathematics, and reading scores, although for some countries test scores are available only for the 1990s. In this study, Barro found that both the quantity of schooling and the quality as measured by science scores have an effect on growth, with quality being more important than quantity. Thus far, owing to a lack of good data, the literature has not investigated the effects of changes in school quality over time.

#### Reverse Causality and Omitted Variables

The authors of most of the recent studies have noted that education seems to have an implausibly large effect on economic growth (in addition to the cited studies, see Bils and Klenow 2000). This may be a result of either reverse causality or omitted variables. Reverse causality occurs as individuals anticipate that higher societal school enrollment will lead to greater economic growth. This causes them to anticipate greater wage gains from investments in education, which in turn affects their schooling decision. As Hanushek and Kimko argued, however, reverse causality is less plausible for the quality of schooling than for the quantity. The thornier problem is omitted variables: Countries that are committed to higher economic growth are likely to undertake a range of pro-growth policies, some of which may be hard to quantify. The education variable will pick up the effect of these other policies. More fundamentally, as Krueger and Lindahl pointed out, macroeconomic studies tend to treat schooling decisions as exogenous; they do not investigate why students in some countries enroll more in school, or learn more from school, than students in other countries.

# Appendix B: Simulations of Changes in Educational Attainment and Returns to Education

Appendix Table 1 shows the simulations of changing educational attainment and earnings for given levels of educational attainment for black men, black women, Hispanic men, and Hispanic women. Separate simulations were performed using 1979–1980 and 1999–2000 worker characteristics and Bradbury's "lower-bound" and "upper-bound" regressions. Appendix Table 2 presents the analogous simulations for women versus men.

In the simulations examining the wage gap between the sexes, men and women were given the mean characteristics of both sexes for all explanatory characteristics except for educational attainment. However, the coefficients—which represent the effects of these characteristics on wages—were taken from Bradbury's regressions looking at only the male population.

## Appendix Table 1

Results of Simulation Exercises Regarding the Racial and Ethnic Wage Gaps

	Simulations					
	Educational attainment mix and the returns to education for each group, equalizing all other explanatory variables		Each g educa attainm and th minority to edu	group's ational ient mix ie non- i returns ucation	Non-minority educational attainment mix and each group's returns to education	
Actual weekly earnings	lower- bound	upper- bound	lower- bound	upper- bound	lower- bound	upper- bound
478.44 627.25 148.81 380.14 408.07 27.93 474.79 622.38 147.59 353.81 407.63 53.82	484.96 614.46 129.50 368.98 404.00 35.02 393.25 617.18 223.93 342.50 402.32 59.82	461.11 617.31 156.20 371.34 406.27 34.93 381.77 618.09 236.33 330.21 403.52 73.31	582.71 614.46 31.74 395.87 404.00 8.13 567.41 617.18 49.77 385.51 402.32 16.81	575.64 617.31 41.68 391.57 406.27 14.70 553.60 618.09 64.49 370.51 403.52 33.01	513.23 614.46 101.22 376.97 404.00 27.03 442.98 617.18 174.21 368.77 402.32 33.55	497.71 617.31 119.61 386.72 406.27 19.55 445.33 618.09 172.77 371.06 403.52 32.46
496.72 634.86 138.13 425.47 488.41 62.94 420.90 658.61 237.72 363.83 495.29 131.45	490.19 613.55 123.37 414.33 482.44 68.11 382.40 643.82 261.43 379.09 484.73 105.64	466.33 615.24 148.91 425.25 483.32 58.07 364.24 646.97 282.73 356.60 488.38 131.77	586.22 613.55 27.33 463.54 482.44 18.89 543.29 643.82 100.53 427.77 484.73 56.96	581.13 615.24 34.10 457.78 483.32 25.55 525.46 646.97 121.51 408.12 488.38 80.26	510.85 613.55 102.70 430.51 482.44 51.93 464.33 643.82 179.50 432.65 484.73 52.08	492.42 615.24 122.82 449.55 483.32 33.77 465.19 646.97 181.78 434.09 488.38 54.28
	Actual weekly earnings 478.44 627.25 148.81 380.14 408.07 27.93 474.79 622.38 147.59 353.81 407.63 53.82 496.72 634.86 138.13 425.47 488.41 62.94 420.90 658.61 237.72 363.83 495.29 131.45	Educa attainm and the to educ each a equali otti explai varia       Actual weekly     lower- bound       478.44     484.96       627.25     614.46       148.81     129.50       380.14     368.98       408.07     404.00       27.93     35.02       474.79     393.25       622.38     617.18       147.59     223.93       353.81     342.50       407.63     402.32       53.82     59.82       496.72     490.19       634.86     613.55       138.13     123.37       425.47     414.33       488.41     482.44       62.94     68.11       420.90     382.40       658.61     643.82       237.72     261.43       363.83     379.09       495.29     484.73       131.45     105.64	Educational attainment mix and the returns to education for each group, equalizing all other explanatory variables       Actual weekly     upper- bound       weekly     lower- bound     upper- bound       478.44     484.96     461.11       627.25     614.46     617.31       148.81     129.50     156.20       380.14     368.98     371.34       408.07     404.00     406.27       27.93     35.02     34.93       474.79     393.25     381.77       622.38     617.18     618.09       147.59     223.93     236.33       353.81     342.50     330.21       407.63     402.32     403.52       53.82     59.82     73.31       496.72     490.19     466.33       634.86     613.55     615.24       138.13     123.37     148.91       425.47     414.33     425.25       488.41     482.44     483.32       62.94     68.11     58.07       420.90     382.40     364.24	Simulation       Educational attainment mix and the returns to education for each group, equalizing all other explanatory variables     Each g education attainment and the explanatory variables       Actual weekly     lower- bound     upper- bound     lower- bound       478.44     484.96     461.11     582.71       627.25     614.46     617.31     614.46       148.81     129.50     156.20     31.74       380.14     368.98     371.34     395.87       408.07     404.00     406.27     404.00       27.93     35.02     34.93     8.13       474.79     393.25     381.77     567.41       622.38     617.18     618.09     617.18       147.59     223.93     236.33     49.77       353.81     342.50     330.21     385.51       407.63     402.32     403.52     402.32       53.82     59.82     73.31     16.81       496.72     490.19     466.33     586.22       634.86     613.55     615.24     613.55       138.131 <td>Educational attainment mix and the returns to education for each group, equalizing all other explanatory variables     Each group's educational attainment mix and the non- minority returns to education       Actual weekly earnings     lower- bound     upper- bound     lower- bound     lower- bound     upper- bound     lower- bound</td> <td>Educational attainment mix and the returns to education for each group, equalizing all other     Each group's educational attainment mix and the ron- minority returns to education     Non-m educational attainment mix and the non- minority returns to education       Actual weekly     lower- bound     upper- bound     lower- bound     upper- bound     lower- bound     lower- bound</td>	Educational attainment mix and the returns to education for each group, equalizing all other explanatory variables     Each group's educational attainment mix and the non- minority returns to education       Actual weekly earnings     lower- bound     upper- bound     lower- bound     lower- bound     upper- bound     lower- bound	Educational attainment mix and the returns to education for each group, equalizing all other     Each group's educational attainment mix and the ron- minority returns to education     Non-m educational attainment mix and the non- minority returns to education       Actual weekly     lower- bound     upper- bound     lower- bound     upper- bound     lower- bound     lower- bound

Appendix Table 2 Results of Simulation Exercises Regarding the Wage Gaps by Sex

				Simu	ations			
	Actual	Educa attainm and the r educatior group, e all c explar varia	Educational attainment mix and the returns to education for each group, equalizing all other explanatory variables		Each group's educational attainment mix and the male return to education		Male educational attainment mix and each group's returns to education	
Panel A 1979–1980	weekly earnings	lower- bound	upper- bound	lower- bound	upper- bound	lower- bound	upper- bound	
Women Men <i>Difference</i>	404.62 612.15 <i>207.53</i>	428.07 567.79 <i>139.72</i>	424.20 583.65 <i>159.45</i>	570.88 567.79 <i>-3.0</i> 8	586.92 583.65 <i>-3.26</i>	429.03 567.79 <i>138.7</i> 6	424.31 583.65 <i>159.34</i>	
Panel B 1999–2000								
Women Men <i>Difference</i>	479.41 620.44 <i>141.03</i>	489.67 579.82 <i>90.15</i>	501.29 587.43 <i>86.13</i>	590.39 579.82 <i>10.58</i>	600.71 587.43 	481.55 579.82 <i>98.27</i>	488.73 587.43 <i>98.70</i>	
Source: Author's	s estimates an	d Bradbury (2	2002).					

# Discussion

# EDUCATIONAL ATTAINMENT AS A CONSTRAINT ON ECONOMIC GROWTH AND SOCIAL PROGRESS

Lawrence F. Katz\*

Yolanda Kodrzycki has produced an insightful and informative overview carefully documenting recent trends in U.S. educational attainment, examining differences in educational outcomes by demographic groups, and exploring the implications of these patterns for economic growth and inequality. She shows that the overall educational attainment of the U.S. adult population (measured by years of schooling or by high school and college degrees) increased substantially from 1970 to 2000, but the rate of progress has been rather slow since the mid-1970s for successive cohorts of new labor market entrants. This pattern (along with other demographic trends such as the aging of the workforce) suggests slower growth in the educational attainment of the U.S. labor force in future decades.

Kodrzycki also documents the persistence of substantial differences in completed schooling by race and ethnicity with little narrowing of the large white–black and white–Hispanic gaps in college-completion rates for younger cohorts over the past 25 years. Furthermore, large racial and ethnic gaps in wages and in a measure of academic achievement (as proxied by average literacy proficiency scores) are apparent for adult U.S. workers, even when conditional on the level of completed schooling. Additionally, substantial racial and ethnic differences in academic achievement (as measured by standardized reading test scores) and differences in access to computers remain for current cohorts of U.S. students. She interprets these group differences in earnings and academic achievement within completed schooling groups as reflecting differences in schooling quality and returns to education by race and ethnicity. She

<sup>\*</sup>Professor of Economics, Harvard University.

concludes that policies to raise the quality of schooling and the labor market returns to schooling for minority groups are crucial for reducing U.S. social inequities and, especially given the shifting demographics of the U.S. workforce, could be important for improving U.S. economic growth prospects.

Since I largely agree with Kodrzycki's thoughtful summary of the trends, I would like to focus on just a few issues. First, I would like to place the recent slowdown of the rate of growth of U.S. educational attainment into historical perspective and sketch some of the implications for wage inequality and economic growth. Second, I will discuss the role of high and rising residential segregation by economic status for educational policies and outcomes. And I will briefly mention some issues related to Kodrzycki's conclusion that differing returns to education are the key factor behind U.S. racial and ethnic wage differences.

# **RECENT CHANGES IN HISTORICAL PERSPECTIVE**

Disparities in the economic fortunes of American families have increased significantly over the past 25 years. Economic inequality in terms of wages, family income, and wealth expanded rapidly in the 1980s and early 1990s, reaching higher levels in the mid-1990s than in any time in (at least) the past 60 years. The strong economic boom of the late 1990s led to substantial real-wage and income growth for low-income families and even narrowed wage dispersion in the bottom half of the distribution. But U.S. wage and income inequality remains much higher today than prior to the 1980s and much higher than in other advanced economies (Katz and Autor 1999; Mishel, Bernstein, and Schmitt 2001). Labor market changes that have greatly increased overall wage dispersion and shifted wage and employment opportunities in favor of the more educated and the more skilled have played an integral role in this process.

The rising inequality and educational wage differentials of the last 25 years represent a break from the pattern of most of the twentieth century. Most of the century was a "human capital" century in which the United States moved ahead of the world in educational attainment, first through the "high school movement" of the first half of the twentieth century and then with the expansion of college education following World War II (Goldin 2001; Goldin and Katz 2001a). The rapid expansion of educational attainment was associated with great technological dynamism, rapid economic growth, declining or stable wage inequality, and contained educational wage differentials as rapid skill-supply growth kept pace with rapid skill-demand growth from skill-biased technological change (Goldin and Katz 2001b). But educational wage differentials and overall wage inequality increased sharply in the 1980s through the early 1990s, with some slowing in the second half of the 1990s.

A simple labor market framework emphasizing the role of supply factors, demand factors, and labor market institutions goes reasonably far toward explaining the historical evolution of U.S. educational wage differentials (Katz and Autor 1999). Much evidence shows that new technologies and shifts in the industrial and occupational composition of employment have been skill-biased (education-biased) throughout the twentieth century. But this growth in the relative demand for skill (human capital) was more than matched by rapid growth in the relative supply of skills (educational upgrading) throughout most of the century. Something changed with a sharp slowdown in the growth of educational attainment for U.S. cohorts starting with the baby boom cohorts of the late 1940s and early 1950s. The combination of the slowdown of educational progress across successive cohorts of labor market entrants and shifting demographics (for instance, the aging of the baby boom cohorts and the labor market entrance of smaller baby bust cohorts) has meant a sharp reduction in the growth rate of the relative supply of skills (for example, the relative supply of college-equivalent workers) in the last two decades relative to previous decades. Institutional factors (the erosion of the real value of the minimum wage and of union strength) and weak macroeconomic conditions also contributed to rising wage inequality in the early 1980s, while a boost in the minimum wage and tight labor markets helped to narrow wage inequality from the mid- to late 1990s.

Figure 1 illustrates the slowdown of the rate of increase of educational attainment of U.S. birth cohorts starting with cohorts born around 1950. Average educational attainment increased by 0.08 year per birth cohort (or two full years of schooling for every 25 successive cohorts) for the birth cohorts of 1876 to 1950. But over the last 25 years (the 1950 to 1975 birth cohorts), the educational attainment of young cohorts increased by only 0.68 year (or 0.027 year per cohort). Similar patterns of slowdown hold for the share of workers going to college or graduating from college starting in the 1970s (around the 1950 birth cohort), with some increase in the rate of growth of college completion for the most recent cohorts. The consequence has been that the educational productivity of the U.S. workforce (measured by educational attainment, weighted by educational wage differentials), which expanded by 0.55 percent per year from 1940 to 1980 (and by over 0.60 percent per year in the 1960s and 1970s), slowed down to only 0.35 percent per year for 1980 to 2000 (Goldin and Katz 2001a; DeLong, Goldin, and Katz 2003). The slower growth of the educational attainment of the workforce directly reduces economic growth by slowing the growth in labor force quality and may adversely impact the rate of technological advance. And changes in the growth of the relative supply of skills have a major impact on wage inequality.

In particular, a slowdown in educational expansion, combined with even stable (not declining) growth in the relative demand for more-



educated workers, can generate an increase in educational wage differentials and overall wage inequality. In the United States, the growth of the supply of college-equivalent workers relative to high-school equivalent workers slowed from a rate of 3.8 percent per year from 1960 to 1980 to under 2.5 percent per year in the 1980s and 1990s (Katz and Autor 1999). Countries with decelerations in the rate of educational advance in recent cohorts (United States, United Kingdom, and Canada) have all experienced substantial increases in educational wage differentials, especially for younger cohorts (Card and Lemieux 2001). Countries with continued rapid expansions of educational attainment (France, Netherlands, and Germany) have not experienced similar large increases in educational wage differentials. Slower growth in the relative supply of college-equivalent workers combined with rapid growth in the demand for more-educated workers, partially driven by computerization and related technological and organizational changes, has been a recipe for rising educational wage differentials and wage inequality.

The slowdown in U.S. college enrollment and completion rates has been concentrated among individuals from lower-income and minority families (Ellwood and Kane 2000). Much of the early slowdown might have reflected strained schooling resources from the large baby boom cohorts born in the 1950 and early 1960s, reduced male college-bound rates from the abnormally high levels associated with Vietnam draftavoidance behavior in the late 1960s, and a response to the decline in the college wage premium observed during the 1970s. The large and growing college wage premium of the 1980s and 1990s led to a substantial increase in college-enrollment rates for middle-class youth but not much increase for lower-income youth.

What accounts for the large and growing gaps in college-enrollment rates for youths by parental income? A large share of the differences in college enrollment by family income is driven by differences in academic investments earlier in the life-cycle arising from family inputs, neighborhood influences, and the quality of preschools, primary, and secondary schools (Heckman and Lochner 2000). But substantial differences in college enrollment (and persistence) remain by family income, even when controlling for achievement test scores and high school grades (Ellwood and Kane 2000). This suggests that financing constraints may remain a significant barrier to college for many low- and moderate-income vouths. Much evidence suggests that college-enrollment rates respond to visible changes in college costs for low-income youth (Dynarski 2002). Recent estimates of the rates of return to schooling using quasi-experimental variation in access to college and college costs systematically generate high rates of return to schooling to the marginal (typically low-income) families affected by such policy interventions (Card 1999). This evidence suggests that financing and information barriers remain substantial for some families. It also suggests that improved college financial aid, earlier mentoring policies, and a more transparent financial aid application and information system could have substantial positive payoffs for disadvantaged youth and could feed back into secondary school performance by creating better incentives for high academic achievement.

# GROWING RESIDENTIAL SEGREGATION BY ECONOMIC STATUS

Poverty in the United States has become increasingly concentrated in inner cities. Table 1 shows that poverty rates in suburban and nonmetropolitan areas of the United States declined substantially over the past 40 years, but poverty persisted in central-city areas. The share of the poor in central cities increased from 27 percent in 1959 to 42 percent in 2000 despite growing suburbanization that reduced the share of the population in central cities. A broader pattern of growing residential segregation by economic status (family income) is also apparent in U.S. census data since 1970 (Watson 2002). The growth of income inequality itself plays an important role in increasing residential segregation by economic status as wealthier families increasingly can outbid poorer families for neighborhood amenities.

The growing concentration of poverty in inner cities has potentially disturbing implications because of evidence that residential neighbor-

Table 1 The Growing Concentration of U.S. Poverty in Central Cities, 1959–2000						
	Poverty Rates (i	n Percent) by Residence, 1	959, 1973, 1994, and	2000		
	Overall	Central City	Suburbs	Non-Metro		
1959	22.4	18.3	12.2	33.2		
1973	11.1	14.0	6.4	14.0		
1994	14.5	20.9	10.3	16.0		
2000	11.3	16.1	7.8	13.4		
	Percentage of	the Total Population and o	f the Poor in Central C	lities		
	All	Poor				
1959	32.2	26.9				
1973	29.6	37.4				
1994	29.4	42.2				
2000	29.1	41.6				
Source: L	S Census Bureau	Historical Poverty Tables: Pe	onle Tables 2 and 8	13 February 2002		

<www.census.gov/hhes/poverty/histpov/perindex.html>.

hoods are associated with the current well-being and future opportunities of residents. Children who grow up in poor neighborhoods fare substantially worse on a wide variety of outcomes than those who grow up with more affluent neighbors. One interpretation of these findings is that residential location greatly affects access to opportunity through peer influences on youth behavior and through substantial observed differences by neighborhood wealth-such as school quality, safety from crime, and supervised after-school activities. Although attempts to sort out the true causal impacts of neighborhoods on the labor market prospects of minority and disadvantaged children from other (hard-toobserve) family background factors are fraught with difficulties, recent work on the quasi-experimental Gautreaux and random-assignment Moving to Opportunity housing mobility programs indicate that moves from high-poverty, inner-city areas to lower-poverty areas can have large positive impacts on children's human-capital development, including educational attainment, test scores, health, and measures of problem behaviors (Katz, Kling, and Liebman 2001; Ludwig, Ladd, and Duncan 2001; Rosenbaum 1995).

Changes in the residential concentration of poverty may greatly impact the ability of schools to deal with social problems and disadvantages. School policies need to be understood in this context. And housing mobility policies (housing vouchers) may be an important complement to educational policies in improving human capital development. Furthermore, the success of residentially based job training programs for disadvantaged youths (for example, the Job Corps) relative to similar training programs without a residential component is further evidence of the need for taking peer and neighborhood interactions into account in the design of education and training programs (Krueger 2002).

# DECOMPOSING RACIAL AND ETHNIC WAGE DIFFERENTIALS

Finally, I have a small quibble with Kodrzycki's analysis of the role of differential returns to education as a source of white-black and white-Hispanic wage gaps. She presents simulations that compare the impacts on racial and ethnic wage differentials of raising minority educational attainment to the same level as whites' (given observed estimated returns to education for the minority group) and of giving the minority group the white returns to education (holding minority educational attainment constant). She concludes that equalizing returns to education would go much further towards reducing racial and ethnic wage differentials than equalizing educational attainment. But the simulation she actually performs appears to involve not the equalization of rates of returns to schooling but the equalization of wages themselves within education groups. In other words, Kodrzycki correctly observes that the majority of white-black wage differentials occurs within education groups. But, in fact, the estimates of returns to schooling by race from Bradbury (2002) are not that different for whites and blacks for recent years. And the equalization of minority-white differences in these estimated returns to education themselves would have only a modest impact on minority-white earnings differences.

For example, using data from the 1999–2000 Current Population Survey outgoing rotation groups for full-time workers, I find that equalizing white–black returns to education reduces the white–black weekly wage differential only by 4 (2) percentage points for nonelderly adult males (females) and by even a smaller amount for younger cohorts. The equalization of educational attainment by race actually has a somewhat larger impact on racial wage differentials (typically 6 percentage points) for the groups I examined.

On the other hand, Kodrzycki's paper and simulations do make the important point that the racial and ethnic wage differentials are quite substantial in the United States even when looking at individuals with the same years of completed schooling. Neal and Johnson (1996) and others suggest that gaps in academic achievement related to school quality, neighborhood, and family backgrounds play a large role in these wage differentials for younger cohorts. Although much evidence suggests direct racial discrimination still plays a role in the U.S. labor market (Altonji and Blank 1999), much of the remaining racial and ethnic gap may relate to family, neighborhood, and school resources (development deficits), and to lingering racial stigmas as emphasized by Loury (2002).

### References

- Altonji, Joseph G. and Rebecca Blank. 1999. "Race and Gender in the Labor Market." In Handbook of Labor Economics, edited by O. Ashenfelter and D. Card, vol. 3C, 3143-3259. Amsterdam: Elsevier.
- Bradbury, Katharine L. 2002. "Education and Wages in the 1980s and 1990s: Are All Groups Moving Together?" New England Economic Review Q1: 19-46.
- Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, edited by O. Ashenfelter and D. Card, vol. 3A, 1801-63. Amsterdam: Elsevier.
- Card, David and Thomas Lemieux. 2001. "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis." *Quarterly Journal of Economics* 116 (2): 705-46.
- DeLong, J. Bradford, Claudia Goldin, and Lawrence F. Katz. 2003. "Sustaining U.S. Economic Growth." In Agenda for the Nation, edited by H. Aaron, J. Lindsay, and P. Nivola. Washington, DC: Brookings Institution, forthcoming.
- Dynarski, Susan. 2002. "The Behavioral and Distributional Implications of Aid for College." American Economic Review 92 (2): 279-85.
- Ellwood, David T. and Thomas J. Kane. 2000. "Who Is Getting a College Education? Family Background and the Growing Gaps in Enrollment." In *Securing the Future*, edited by S. Danziger and J. Waldfogel, 283-324. New York: Russell Sage Foundation.
- Goldin, Claudia. 2001. "The Human-Capital Century and American Economic Leadership: Virtues of the Past." *Journal of Economic History* 61 (2): 263-92.
- Goldin, Claudia and Lawrence F. Katz. 2001a. "The Legacy of U.S. Educational Leadership: Notes on Distribution and Economic Growth in the 20<sup>th</sup> Century." American Economic Review 91 (2): 18-23.
- Heckman, James J. and Lance Lochner. 2000. "Rethinking Education and Training Policy: Understanding the Sources of Skill Formation in a Modern Economy." In *Securing the Future*, edited by S. Danziger and J. Waldfogel, 47-83. New York: Russell Sage Foundation.
- Katz, Lawrence F. and David H. Autor. 1999. "Changes in the Wage Structure and Earnings Inequality." In *Handbook of Labor Economics*, edited by O. Ashenfelter and D. Card, vol. 3A, 1463-1555. Amsterdam: Elsevier.
- Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics* 116 (2): 607-54.
- Krueger, Alan B. 2002. "Inequality: Too Much of a Good Thing." Princeton University, unpublished paper (April).
- Loury, Glenn C. 2002. The Anatomy of Racial Inequality. Cambridge, MA: Harvard University Press.
- Ludwig, Jens, Helen F. Ladd, and Gregory J. Duncan 2001. "The Effects of Urban Poverty on Educational Outcomes." *Brookings-Wharton Papers on Urban Affairs* 2001: 47-201.
- Mishel, Lawrence, Jared Bernstein, and John Schmitt. 2001. The State of Working America, 2000-01. Ithaca, NY: Cornell University Press.
- Neal, Derek and William R. Johnson. 1996. "The Role of Pre-Market Factors in Black-White Wage Differences." Journal of Political Economy 104 (5): 869-95.
- Rosenbaum, James E. 1995. "Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program." *Housing Policy Debate* 6: 231-69.
- Watson, Tara. 2002. "Inequality and the Rising Income Segregation of American Neighborhoods." Harvard University, unpublished paper (June).

# Discussion

# Educational Attainment as a Constraint on Economic Growth and Social Progress

# Paulo Renato Souza\*

Brazil offers an intriguing case where education is not related to development. Between 1900 and 1975, Brazil had the second highest rate of growth in the world, second only to Japan, though rates have slowed since then. Brazil also had one of the worst income distributions in the world. Only a few small African countries had a more unequal distribution than we did. In 1960, Brazil had an illiteracy rate of 40 percent, and 40 percent of our children between the ages of 7 and 14 were not in school, suggesting that development, or growth at least, was not related to education, though income distribution probably was.

The situation was still very difficult in 1994, when I took charge of the Ministry of Education. I have been in the position for seven and a half years, but I am an economist by training. I became involved in education administration in 1984 when I was appointed Secretary of Education of the state of San Paolo, which is the largest state in the country with the largest system of education. I was appointed because I taught labor economics, and there was a huge teachers' strike. I was hired to end the strike and to resolve the labor problems. After that, I remained in education as President of the State University of Campinas. Then I was Operations Manager at the Inter-American Development Bank for four years, but returned as the Minister of Education.

Brazil has a huge educational system. We have 60 million students in the entire system. Almost one-third, or 30 percent, of our population is attending some kind of school. But even in 1994 we still had only 87 percent of our children between the ages of 7 and 14 attending school; about 13 percent were not in school, meaning that, by the end of the

<sup>\*</sup>Minister of Education, Ministry of Education, Brazil.

twentieth century, we still did not have universal access to elementary education, which the United States had by the end of the nineteenth century.

At that time, 25 percent of the children in the northeast of Brazil, which is the poorest area of the country, were not in school. Nationwide, 25 percent of the children in the lowest income quintile were not in school. Twenty percent of black children were not in school. The illiteracy rate for those aged 15 to 19 years was 6.8 percent nationwide. But in the northeast it was 16 percent. And 17 percent of the population aged 15 and over was illiterate. Just 50 percent of those who began primary school finished it. And the average time taken to complete the eight years of primary school was 12 years, because so many children failed their yearend exams and had to repeat grades. Extensive grade repetition does not occur in the United States or in England, but in Brazil, and in Latin America in general, it is a huge problem.

Because of grade repetition, at the end of elementary school, having spent 12 years in school already, most children went into the labor market instead of continuing on to high school. As a result, any given member of the Brazilian workforce had an average of only 5.3 years of schooling. The proportion of each cohort that attended high school or university was half that of neighboring countries like Argentina, Chile, and Mexico, and nowhere near that of the United States or Europe. However, we had the best higher-education system in Latin America, with high quality especially in the public universities and with quite a diversified and sophisticated system of research and graduate studies. Our education system had huge contrasts then.

This model of education was compatible with the development Brazil experienced. There were plentiful natural resources and cheap, unskilled labor. At the same time, Brazil was also able to make some very important contributions to technology in the production of oil, in communications, and in agriculture. Some areas of educational excellence were essential to develop such technologies.

Of course, the world has changed. We are now living in the age of knowledge. We are living in a new technological revolution. This new society makes some new demands on our educational system. Today it is necessary to learn how to learn. It is no longer acceptable to concentrate education in just one period of our lives. To exercise citizenship in any aspect, it is necessary to keep learning our whole lives.

Developing the ability to learn has become the primary goal of education. It is necessary to universalize access to basic education, including preschool, elementary school, and secondary school, and it is no longer useful to think in terms of the transmission of knowledge. Rather, basic education should develop the ability to reason, to learn, to understand, and to criticize. This distinction is important because instead
of thinking about years of schooling, we have to think about the content of our basic education in a way that is different from the past.

On the other end, we also have to offer more opportunities for lifelong education, which means increasing vocational training, expanding higher education, encouraging the introduction of new teaching technologies and distance education, and making the structure of secondary education more flexible, in order to permit more frequent entry and exit from the system. These were the challenges, on top of the problem of providing more universal access, that confronted our system in 1994.

I am not going to discuss the main policies we developed,<sup>1</sup> but I would like to point out some recent results. Between 1994 and 2000, we were able to increase the overall enrollment rate in elementary schools from 87 percent to 96.3 percent. We estimate that enrollment is now around 97 percent. We dramatically reduced the difference in enrollment rates across income quintiles: Between 1992 and 1999, the enrollment rate in the lowest quintile increased from 75 percent to 93 percent, approaching the highest quintile's rate of 99 percent. Differences between races have also diminished: Indigenous enrollment rose from 77 percent to 87 percent, and black enrollment rose from 79 percent to 93 percent.

We also reduced grade-repetition rates. The enrollment level in elementary schools grew by only 11 percent in this period, despite the fact that the enrollment rate among children considered to be of elementaryschool age grew from 87 percent to 97 percent, and despite the fact that the size of this age group increased. This is because we substantially improved grade-to-grade approval rates and reduced repetition rates. Between 1994 and 2000, completions at elementary schools grew by 67 percent. Enrollment in high schools rose 71 percent and completions 102 percent. In these six years, total university enrollment grew by 62 percent, indicating that, with a big effort and a huge fiscal reform, we are now catching up.

With a strong decentralization policy and with community participation in the administration of the system and in the schools themselves, we are now in a position to face the challenge of having education enhance development and reduce disparity in Brazil. Of course, we still have to continue to keep an eye on improving quality. Overall, the Brazilian experience offers an interesting contrast to the situation in the United States.

<sup>&</sup>lt;sup>1</sup> For more information see Paulo Renato Souza, *Education and Development in Brazil*, Ministry of Education: October 2000.

### Social and Nonmarket Benefits from Education in an Advanced Economy

#### Barbara L. Wolfe and Robert H. Haveman\*

The extent to which human capital, especially schooling, contributes to social well-being and economic growth is an important question, and has been addressed in numerous research studies. The results of these studies are diverse, and hence controversial and widely debated. Evidence on this issue has important implications for public policies toward education and the optimal public/private balance in the financing of educational services.

One important line of research builds on the "human capital model" developed by Becker (1964) and Mincer (1962). Here the strategy has been to empirically estimate the returns to incremental schooling largely in the form of market-valued increases in productivity associated with more schooling. The value of this increase in skills and productivity is reflected in earnings differences between identical individuals with different levels of schooling. The 40 years of research on this question has been voluminous, and only recently has a consensus emerged regarding the wage returns to schooling.<sup>1</sup>

A second important approach is often referred to as "endogenous growth" analysis (Barro 1997, 2001 reviews this literature). The theoretical

<sup>\*</sup>Professor of Economics, Population Health Sciences, and Public Affairs; and Professor of Economics and Public Affairs, respectively, University of Wisconsin–Madison. The authors wish to thank Samuel Zuvekas, Cullen Goretzke, and Elise Gould for their contributions to this paper, John Wolf for editorial assistance, and John Ermisch for running additional estimates for us. Any views expressed are those of the authors alone.

<sup>&</sup>lt;sup>1</sup> An important issue in this literature concerns the potential upward bias in estimates of the return to schooling caused by an "ability bias." This was an important topic in an early review of this literature, by Griliches (1977). The current consensus is discussed by Card (2001), who concludes that estimates based on instrumental variables seem to be higher than the earlier studies based on Mincerian wage functions.

growth models of this genre view economic growth as dependent on purposeful research leading to new and improved products and ways of producing, which are then spread across sectors and nations. Empirical studies estimating this process use cross-national data on per capita output (gross domestic product per person), levels of physical and human capital, and characterizations of the demographics of the national population and the level and quality of its policies and institutions. They seek to reveal the persistent effect of policy levels and change on the growth rate of per capita output. Education quantity and quality, often characterized by measures of average years of school attainment of gender/age groups (see Barro and Lee forthcoming) and by test score indicators, are typically one of the central variables of interest. Again, the focus is on the determinants of market-based outcomes. The findings from this literature are diverse and controversial.

In this paper, we emphasize that a full evaluation of the effect of schooling on social well-being requires that we move beyond these market-based effects of education. As we show, the list of the potential effects of schooling that are not reflected in estimates of market returns is extensive, and involves both nonmarket effects that are private (in the sense of being captured by individuals) and social effects involving the public goods or "spillover" effects of schooling. We argue that these effects may be large, and under certain assumptions may be as large as the market-based effects of education. Irrespective of their magnitude, these effects are relevant for determining the optimal level of social (and public-sector) investment in schooling. We first catalog the market, private nonmarket, and social outcomes of schooling, and cite some of the more important contributions to the research literature providing evidence of such impacts. We cite papers that emphasize market-based and market-valued contributions of education in both the "rate of return" and "growth" literatures, and concentrate on the studies that attempt to assess the private nonmarket and social effects of schooling.<sup>2</sup> Our review includes studies that use data from developing countries.

As we have indicated, the catalog of private nonmarket and social effects of education is long, and includes such relationships as these:

- a likely positive link between one's own schooling and the schooling received by one's children;
- a likely positive association between one's own schooling and the health status of one's family members;
- a likely positive relationship between one's own education and one's own health status;

<sup>&</sup>lt;sup>2</sup> Prior studies that also address this question are Haveman and Wolfe (1984, 2001), Michael (1982), McMahon (2000), Greenwood (1997), and Wolfe and Zuvekas (1997). Behrman and Stacey (1997) discuss a variety of sources of these nonmarket effects.

- a likely positive relationship between one's own education and the efficiency of choices made, such as consumer choices (the efficiency of which contributes to a well-being similar to the contribution of money income);
- a relationship between one's own schooling and fertility choices (for example, decisions of one's female teenage children regarding nonmarital childbearing); and
- a relationship between schooling in one's neighborhood and youth decisions regarding their level of schooling, nonmarital childbearing, and participation in criminal activities.

After presenting the catalog of social and private nonmarket effects of schooling and the results of the literature that provides evidence on the extent of these effects, we propose a method for valuing the private nonmarket effects of schooling under a set of demanding assumptions. Then, we present some illustrative estimates of these values, recognizing the heroic nature of the assumptions on which they rest. We find that the values of these private nonmarket returns to education are potentially very large, though as yet not accurately estimated. We conclude by noting that evaluating both the optimal level of social investment in education and the public/private balance in financing education requires a comprehensive assessment of all of the returns to schooling—market, nonmarket, and external/public goods effects.

#### THE SIZE OF THE EDUCATION SECTOR: SOME BACKGROUND

The education sector in western developed countries (in particular, the OECD countries) is very large, and substantial financial contributions by governments and private citizens are required to cover the total costs of providing schooling services. In all of these countries, the bulk of the social costs of providing schooling services at the elementary and secondary level are borne by taxpayers, as most schools are publicly funded. Among the set of countries shown in Table 1, 1997 spending for primary education is highest in Denmark-\$6,913 per student (in 1997 dollars, translated by the OECD purchasing power parity index)-with Austria and Switzerland close behind. At the secondary level, Belgium records the highest per student expenditure at \$9,111, followed by Austria. Switzerland, the United States, France, and Denmark all spent more than \$7,000 per secondary student in 1997. At the higher education level, Japan had the highest expenditures per pupil at \$18,914, followed by Switzerland, the United States, and Canada, all with per pupil expenditures greater than \$14,000. Turkey had the lowest per pupil higher education expenditure at about \$2,400.

Table 2 presents a tabulation of the percentage of national GDP

Public Spending per Pupli: Selected Countries, 1997			
Country	Primary	Secondary	Higher Education
Australia	\$3,751	\$5,794	\$11,240
Austria	6,258	8,213	9,993
Belgium <sup>a</sup>	5,205	9,111	_
Canada	_	_	14,816
Czech Republic	1,942	3,643	5,478
Denmark	6,913	7,285	7,294
Finland	4,643	5,009	7,190
France	3,735	7,118	7,058
Germany	3,460	4,536	9,621
Greece	2,351	2,581	3,990
Hungary	2,035	2,093	5,430
Iceland	_	_	_
Ireland	2,571	3,868	8,171
Italy	5,073	6,284	5,972
Japan	5,203	_	18,914
Korea	3,327	3,909	6,227
Luxembourg	_	_	_
Mexico	871	1,651	4,628
Netherlands	_	_	_
New Zealand	_	_	_
Norway	_	4,174	10,108
Poland	1,446	_	4,395
Portugal	3,248	4,264	_
Russia	_	_	_
Spain	3,560	5,386	5,335
Sweden	5,520	5,429	12,785
Switzerland	6,237	7,243	16,376
Turkey	_	_	2,397
United Kingdom	3,206	4,982	_
United States	5,961	7,462	14,864

Table 1 Public Spending per Pupil: Selected Countries, 1997

Note: Data adjusted to U.S. dollars using the purchasing power parity (PPP) index.

<sup>a</sup> Data for Flemish Belgium only.

Source: Organization for Economic Cooperation and Development (1996, 1998, and 1999) and unpublished data from www.oecd.org.

allocated to all levels of schooling, and for the primary/secondary and higher education components of the education sector. Across the countries, the public sector allocates an average of about 5.1 percent of GDP to the provision of schooling services. This percentage ranges from 3.5 to 3.6 percent in Greece and Japan to 6.5 percent or more in the Nordic countries. On average, about 3.6 percent of GDP is allocated to primary and secondary schooling, and 1 percent of GDP is allocated to higher education. This tabulation of public sector costs, however, understates the full cost of providing educational services, especially at the higher education level. Whereas about 93 percent of the total costs of primary

Table 2

Country	All Institutions <sup>a</sup>	Primary and Secondary <sup>b</sup>	Higher Education
Australia	4.3	3.3	1.0
Austria	6.0	4.2	1.3
Belaium <sup>c</sup>	4.8	3.3	0.8
Canada	5.4	4.0	1.2
Czech Republic	4.5	3.2	0.7
Denmark	6.5	4.3	1.1
Finland	6.3	3.8	1.7
France	5.8	4.1	1.0
Germany	4.5	2.9	1.0
Greece	3.5	2.5	1.0
Hungary	4.5	2.9	0.8
Iceland	5.1	3.9	0.7
Ireland	4.5	3.4	1.0
Italy	4.6	3.4	0.6
Japan	3.6	2.8	0.5
Korea	4.4	3.4	0.5
Luxembourg	4.2	4.1	0.1
Mexico	4.5	3.3	0.8
Netherlands	4.3	2.9	1.1
New Zealand	6.1	4.7	1.0
Norway	6.6	4.4	1.3
Poland	5.8	3.8	1.2
Portugal	5.8	4.4	1.0
Russia	_	_	_
Spain	4.7	3.5	0.9
Sweden	6.8	4.7	1.6
Switzerland	5.4	4.0	1.1
Turkey	_	_	0.8
United Kingdom	4.6	3.4	0.7
United States	5.2	3.5	1.4
Average for Year	5.1	3.6	1.0

Total Public Direct Expenditures on Education as a Percentage of the Gross Domestic Product: Selected Countries, 1997

Note: Direct public expenditure on educational services includes both amounts spent directly by governments to hire educational personnel and to procure other resources, and amounts provided by governments to public or private institutions.

<sup>a</sup> Includes pre-primary and other expenditures not classified by level.

<sup>b</sup> Because of the implementation of a new classification system, post-1996 data are not comparable with earlier data.

 $^{\rm c}$  Data for Flemish Belgium only.

Source: Organization for Economic Cooperation and Development (OECD), Education Database;

Annual National Accounts, vol. 1, 1997; and *Education at a Glance*, 2000. (This table was prepared July 2000.) Data drawn from *Digest of Education Statistics*, 2000 < http://nces.ed.gov/pubs2001/digest/dt412.html>

and secondary schooling are borne by the public sector in the United States,<sup>3</sup> the public sector bears only about 30 percent of the total costs of higher education, defined to include the value of the time spent by students as indicated by the earnings that they forgo by choosing to seek education rather than working and earning. About 20 percent of the total costs (so defined) are borne by parents and students in the form of tuition, fees, books, and supplies. According to one calculation for the United States for the early 1990s, over 50 percent of the full cost of higher education is accounted for by such forgone earnings.<sup>4</sup>

When these privately borne financial and forgone earnings costs are accounted for, the full cost of higher education is about three times the public sector cost indicated in Tables 1 and 2. If the social cost in the form of privately borne costs plus forgone earnings is added to the public sector costs reported in Table 2, the average percentage of GDP allocated to higher education would rise from 1 percent to 3 percent, and the average total social cost of education would increase from 5 percent to 7 percent.

An important question concerns the social benefits attributable to this enormous allocation of resources to the provision of schooling services. Because of the large role of the public sector in financing schooling services, the provision of education services is effectively removed from the market test. As a result, the private benefits of education, as reflected in the willingness to pay of private beneficiaries of schooling services—or, more precisely, the private willingness-to-pay value of incremental schooling—cannot be inferred from market demands, prices, expenditures, and surpluses. Direct measures of these values are required. However, even a full measure of the privately captured benefits from incremental schooling would fail to reflect another source of the gains from education, those in the form of external effects or public goods. These too are important in assessing the full social value of schooling, and like the private gains, these too must be directly measured and assessed.

# MARKET AND NONMARKET EFFECTS OF EDUCATION: A CATALOG

Traditionally, economists have sought to measure the private returns to schooling, in particular, the private market returns reflected in the

<sup>&</sup>lt;sup>3</sup> This large absolute and proportional public expenditure pattern at the elementary and secondary level also exists in most other OECD countries, suggesting a relatively small contribution of private spending in support of schooling services at levels below the higher education level.

<sup>&</sup>lt;sup>4</sup> See Edgmand, Moomaw, and Olson (1994).

effect of additional schooling on earnings.<sup>5</sup> Because not all of the well-being gains that people obtain from schooling are reflected in labor market returns, this approach yields but a partial measure of the full, privately captured returns to education. At least as seriously, it totally neglects the external and public-good-type benefits associated with increased schooling.

In this section, we construct a comprehensive list of the components of the social gains (or losses) associated with education, including private market returns, private nonmarket benefits, and the external and publicgood-type benefits of education. Note that neither of the latter two categories is reflected in the traditional economic estimates of the private returns to schooling reflected in market earnings differences. Our emphasis on these latter two components—the privately captured nonmarket gains and the external/public goods effects of education—reflects our view that a full accounting must consider all of schooling's effects, and not simply those recorded in a single market.

In the Appendix, we present an accounting framework for assessing the social costs and benefits of investments in human capital. The categories of social gain that we distinguish in the following catalog derive from this accounting framework.

In Table 3, we identify a number of categories of market and nonmarket (both private and external/public goods) benefits attributable to schooling, together with a description of the research studies that reflect the magnitudes of these benefits. The studies identified in this table make extensive use of statistical controls for characteristics such as age, race, and other relevant factors in deriving estimates of the magnitude of the effects that are attributable to education.

#### Private Market Returns (Categories 1 and 2)

The first two outcomes categories in Table 3—labor market productivity and nonwage labor market remuneration—capture the gains to education reflected in the traditional "returns to schooling" studies discussed above. Figures 1 and 2 illustrate the simple relationship between schooling and earnings both at a single point in time and over time; the recent increase in the earnings returns to schooling stands out. As of 1998, median earnings of full-time college graduates were \$46,285,

<sup>&</sup>lt;sup>5</sup> The basic equation that is estimated in assessing the private market return to schooling is  $Y_i = \alpha + \beta S_i + \mu_i$ , where  $Y_i$  is earnings of individual i, usually specified in log form,  $S_i$  is the individual's level of schooling,  $\beta$  is the return to schooling parameter to be estimated,  $\alpha$  is the estimated constant term, and  $\mu_i$  is the error term. Added to this are additional control variables to reduce bias from ability and experience.

	Outcome Category	Economic Nature	Existing Research on Magnitude
1.	Individual market productivity	Private; market effects; human capital investment	Extensive research on the magnitude of market earnings (Schultz 1961; Mincer 1962; Hansen 1963; Becker 1964; Conlisk 1971) and of changes over time (Allen 2001). Debate over role of work while acquiring schooling (Light 2001). Analysis exploring approaches to eliminating ability bias and publication bias (Ashenfelter, Harmon, and Oosterbeek 2000).
2.	Nonwage labor market remuneration	Private; market and nonmarket effects	Some research on differences in fringe benefits and working conditions by education level (Duncan 1976; Lucas 1977; Freeman 1981; Smeeding 1983) and wage level (Vanness and Wolfe 2002).
3.	Intrafamily productivity	Private; some external effects; market and nonmarket effects	Relationship between wife's schooling and husband's earnings apart from selectivity is established (Benham 1974). Suggestion that relationship is stronger in entrepreneurial families (Wong 1986) and among those whose spouse is in a skilled position (Neuman and Ziderman 1990). Also, some evidence that own schooling influences spouse's health and decreases mortality (Auster, Leveson, and Sarachek 1969; Grossman 1975; Grossman and Jacobowitz 1981).
4. (	Child quality: level of education and cognitive development	Private; some external effects; market and nonmarket effects	Substantial evidence that a child's education level and cognitive development are positively related to the mother's and father's education (Wachtel 1975; Murnane 1981; Sandefur, McLanahan, and Wojtkiewicz 1989; Dawson 1991; Haveman, Wolfe, and Spaulding 1991; Ribar 1993; Haveman and Wolfe 1994; Duncan 1994; Angrist and Lavy 1996; Ermisch and Francesconi 1997; Smith, Brooks-Gunn, and Klebanov 1997; Lam and Duryea 1999; Duniform, Duncan, and Brooks-Gunn 2000). Extended to a child's self-esteem (Axinn, Duncan, and Thornton 1997). Some evidence that a child's education is positively related to the grandparents' schooling (Blau 1999). Some evidence that education of adults in the neighborhood increases probability of a child's graduating high school (Clark 1992; Duncan 1994; Ginther, Haveman, and Wolfe 2000). Some evidence that increased women's literacy leads to higher human capital of children in developing countries (Behrman et al. 1999).

#### Table 3 Catalog of Outcomes of Schooling

	0	0	
	Outcome Category	Economic Nature	Existing Research on Magnitude
5.	Child quality: health	Private; some external effects	Substantial evidence that child health is positively related to parents' education (Edwards and Grossman 1979; Shakotko, Edwards, and Grossman 1981; Wolfe and Behrman 1982; Behrman and Wolfe 1987; Grossman and Joyce 1989; Strauss 1990; Thomas, Strauss, and Henriques 1991; King and Hill 1993; Glewwe 1999; Lam and Durvea 1999)
6.	Child quality: fertility	Private; some external effects	Consistent evidence that a mother's education is related to a lower probability that daughters will give birth out of wedlock as teens (Antel 1988; Sandefur and McLanahan 1990; Hayward, Grady, and Billy 1992; An, Haveman, and Wolfe 1993; Lam and Duryea 1999; South and Baumer 2000; Haveman, Wolfe, and Wilson 2001).
7.	Own health	Private; modest external effects (Note: Some of the own health benefits from education will be captured in increased earnings, and hence included in <i>category 1</i> )	Considerable evidence that one's own schooling positively affects one's health status (Leigh 1981, 1983; Kemna 1987; Berger and Leigh 1989; Grossman and Joyce 1989; Kenkel 1991; Strauss et al. 1993; Sander 1995); also increases life expectancy (Feldman et al. 1989; King and Hill 1993; Crimmins and Saito 2001); also lowers prevalence of severe mental illness (Robins 1984) including depression (Herzog et al. 1998) and improves ability to deal with stressful events (Thoits 1984) and anger (Schieman 2000). High school graduation lowers mortality rate (Muller 2002). Health advantage of more schooling increases with age (Ross and Wu 1988, 1995).
8.	Consumer choice efficiency	Private; some external effects; nonmarket effects	Some evidence that schooling leads to more efficient consumer activities (Michael 1972; Benham and Benham 1975; Pauly 1980; Rizzo and Zeckhauser 1992; Morton, Zettelmeyer, and Silva-Risso 2001). Home-production schooling may have long-term impacts (Corman 1986). College graduates maintain computational skills over longer period (Pascarella and Terenzini 1991).
9.	Labor market search efficiency	Private; nonmarket effects	Some evidence that costs of job search are reduced and regional mobility increased with more schooling (Metcalf 1973; Greenwood 1975; DaVanzo 1983). Job turnover lower for women with more schooling (Rovalty 1998).
10.	Marital choice efficiency	Private; nonmarket effects	Some limited evidence of improved sorting in marriage market (Becker, Landes, and Michael 1977).

#### Table 3 *(continued)* Catalog of Outcomes of Schooling

	Outcome Category	Economic Nature	Existing Research on Magnitude
11.	Attainment of desired family size	Private	Evidence that contraceptive efficiency is related to schooling (Easterlin 1968; Ryder and Westoff 1971; Michael and Willis 1976; Rosenzweig and Schultz 1989). In developing countries, fertility declines (King and Hill 1993; Lam and Durvea 1999).
12.	Charitable giving	Private and public; nonmarket effects	Some evidence that schooling increases donations of both time and money (Mueller 1978; Dye 1980; Hodgkinson and Weitzman 1988; Freeman 1997).
13.	Savings	Private; some external effects	Controlling for income, some evidence that more schooling is associated with higher savings rates (Solomon 1975).
14.	Technological change	Public	Some evidence that schooling is positively associated with research, development, and diffusion of technology (Nelson 1973; Mansfield 1982; Wozniak 1987; Foster and Rosenzweig 1996). Some evidence that technological change increases returns to those with more education (Bound and Johnson 1992; Autor, Katz, and Krueger 1998; Bartel and Sicherman 1999; Allen 2001).
15.	Social cohesion	Public	Descriptive evidence to suggest that schooling is positively associated with voting (Gintis 1971; Campbell et al. 1976; Wolfinger and Rosenstone 1980; Hauser 2000); with reduced alienation and social inequalities (Comer 1988); with opposition to government repression and reduced support for use of violence in protests (Hall, Rodeghier, and Useem 1986). Suggestion that own education is associated with trust of others and membership in community organizations (Helliwell and Putnam 1990)
16.	Self-reliance or economic independence	Private and public	More education associated with reduced dependence on transfers during prime working years (Antel 1988; Kiefer 1985; Rudd, McKenry, and Nah 1990; An, Haveman, and Wolfe 1993).
17.	Crime reduction	Public	Some evidence that schooling is associated with reduced criminal activity (Yamada, Yamada, and Kang 1991; Ehrlich 1975; Freeman 1995; Lochner and Moretti 2001). Some evidence that education is associated with a reduction in recidivism (Sherman et al. 1998). Some suggestion that quality preschool is associated with a reduction in crime (Reynolds 2000).

#### Table 3 (continued) Catalog of Outcomes of Schoolir



versus \$26,592 for those with only a high school degree. This differential has grown over time in nominal and real terms.<sup>6</sup>

As we have noted, there is an extensive literature estimating the private market returns to schooling. The studies typically make use of individual survey data that include variables describing labor market earnings, the amount of schooling attained, and a variety of other personal characteristics that may affect people's earnings. At a minimum, these other characteristics include gender, age, and often some measure of work experience; the more reliable studies also attempt to control for ability, often by including some assessment of IQ scores or other indicators. The results from these studies vary over time and by the model estimated.

Increasingly, researchers have become concerned with the reliability of these estimates because of the difficulty of controlling for unmeasured and unobserved factors that may both affect earnings and be correlated with the measured variables, especially schooling. If these factors— "ability," "drive," and "family background" come immediately to

<sup>&</sup>lt;sup>6</sup> It should be noted that the returns implied by these comparisons may overstate the true private marketed returns to schooling, as they fail to control for important factors such as ability and experience.



Completed," CRS Report for Congress 95-1081E. Updated October 1, 1999.



mind—are not adequately controlled for, the estimated private market returns to education will be overstated. As an example of the potential for overstating these private market returns to education, Light (2001) concludes that the omission from the estimation of work experience while in school results in estimates of the earnings returns from schooling that are from 4 to 20 percent greater than those found when this factor is statistically controlled for.

A recent literature review by Ashenfelter, Harmon, and Oosterbeek (2000) compares results across several types of studies of the labor market returns to education, distinguished by model, sample, extent of control for relevant variables, and the nature of the labor market (such as country). In their discussion of the potential bias in estimated returns caused by unobserved variables, they focus on the absence of reliable measures of "ability," and hence the difficulty of directly controlling for this trait. They note that researchers have adopted a variety of approaches designed to reduce bias caused by the absence of direct measures of "ability," including the following:

- explicitly including direct measures of ability such as test scores;
- use of data that include information on siblings or twins, so as to control for the common genetic effect (thought to include "ability") among observations; and

• use of instrumental variables representing schooling attainment, generally based on so-called natural experiments.<sup>7</sup>

The first of these approaches is criticized because of the weakness of the variables available to directly measure ability. The second approach also has severe limitations in that individual abilities, apart from their common family source, remain uncontrolled. Those who adopt the third approach (see Card 2001) are open to questions regarding the validity of the instrument chosen for the estimation; in fact, few good instruments seem to be available. All of these approaches are, in addition, subject to the problem of downward-biased estimates of the effect of education on earnings, caused by the presence of measurement error in both schooling and earnings.

Ashenfelter, Harmon, and Oosterbeek then perform a meta-analysis over 27 studies in nine countries of the private market returns to schooling. They find that across these studies, which adopt various approaches to control for unobserved variables, researchers continue to find high private market returns to education. For example, across all of the studies the estimated rate of return to schooling averages 7.9 percent (standard deviation = 0.036). When direct controls for ability are employed, the average return drops to 6.6 percent (standard deviation = 0.026); when data using twins are employed, the average return is 9.2 percent (standard deviation = 0.037); when an instrumental variable approach is employed, the average return is 9.3 percent (standard deviation = 0.041). The authors then adjust for "publication bias" (the tilt inherent in the scholarly publication process leading to a higher probability of acceptance for studies with statistically significant results) and find estimated rates of return from 6.4 to 8.1 percent, with higher rates in this range from studies using an instrumental variables approach.<sup>8</sup> Taken as a whole, estimates of the rate of return are guite consistent and do not change substantially in response to any of the approaches used to adjust for any bias due to unmeasured ability.

<sup>&</sup>lt;sup>7</sup> In this approach, the researcher first estimates the effect of an instrumental variable (one that is believed to be associated with the level of schooling but not labor market earnings) on the level of schooling, and then, as a second step, employs predicted schooling variables from the first stage estimation in a model explaining the level of earnings. See note 1.

<sup>&</sup>lt;sup>8</sup> Comparing the United States to other countries, the authors find rates of return in the United States that are about 1.3 points greater than in other countries (primarily the United Kingdom, which is the source of data in most of the other cases), and they attribute that to the large relative increase in education-related earnings in the United States in recent decades. For example, Ashenfelter and Rouse (1998) find that in the United States the return to an additional year of schooling had grown from 6.2 percent in 1979 to about 10 percent in 1993.

#### Nonmarket Private Returns (Categories 3 through 11)

Intrafamily Effects. Categories 3 through 6 in Table 3 refer to the direct effect on other members of a family when one family member (typically, a parent) is more educated. Consider, for example, the effect of the education of one spouse (say, the wife) on the earnings of the other spouse (the husband) (category 3). The research on this relationship finds a positive and significant effect, suggesting that the information, advice, and assistance in skill acquisition and coping with challenges provided by a more-educated spouse has a larger effect on the other spouse's earnings than the contributions of this sort made by a less-educated spouse. In effect, a spouse's education is a close substitute for a person's own formal education. Studies outside of the United States have explored whether this effect differs by spouse's occupation. Evidence from both Hong Kong and Israel suggests stronger effects for entrepreneurial families and spouses in skilled positions (Wong 1986; Neuman and Ziderman 1990). Some studies also indicate that one's schooling has a positive impact on the health of the spouse.

The educational level of children is clearly tied to the schooling of the parents (category 4). Children of parents who graduate from high school are themselves far more likely to graduate from high school than are children of parents without a high school degree, and parental schooling beyond the high school level increases this probability (Sandefur, McLanahan, and Wojtkiewicz 1989).<sup>9</sup> Similarly, parents with more education tend to have children with a higher level of cognitive development (and "noncognitive" skills<sup>10</sup>), as well as with higher future earnings. There is also evidence of a positive relationship between the educational level of young adults in a community and the probability that children living in the community will complete secondary schooling. Complementing these estimated relationships is recent evidence that grandparents' schooling also is associated with higher levels of children's cognitive development (Blau 1999).

The studies in category 5 suggest that increased schooling of parents, particularly mothers, is also positively associated with higher health status levels of infants and children (as indicated by lower rates of infant

<sup>&</sup>lt;sup>9</sup> The relationship between parental education and a variety of children's attainments is explored in detail below in the section "On Estimating the Value of Nonmarket Impacts of Education."

<sup>&</sup>lt;sup>10</sup> Bowles, Gintis, and Osborne (2001) provide evidence that children's and youths' noncognitive skills (such as attitudes toward risk, ability to adapt to new economic conditions, industriousness, perseverance, and the rate of time preference) are related to future labor market and other indicators of success, and that these noncognitive skills are not captured by measures of cognitive skills. They also suggest that such noncognitive skills and behaviors may be "learned" from parents, and that more and better parental education contributes to children's possession of these skills.

mortality and low birth weight). Similarly, the rate of vaccinations among children is positively related to the educational level of their parents. Evidence for these linkages between parental education and children's health status is also found in studies using data from less-developed countries. The level of parental schooling also seems to be negatively related to the probability that one's child will give birth out of wedlock as a teenager (category 6).

*Own Nonmarket Effects.* Categories 7 through 11 summarize a variety of potential effects of education on one's own well-being that are not captured in labor market performance, and hence are excluded (at least in part) in the estimated privately captured economic returns to schooling.<sup>11</sup>

For the individual, increased schooling appears related to better health and increased life expectancy (category 7). This may be attributable to schooling-related occupational choices (choosing occupations with relatively low occupational hazards), locational choices (electing to live in less-polluted areas), information or skills in acquiring healthrelated information, nutrition and lifestyle (more exercise; less smoking),<sup>12</sup> and/or more appropriate medical-care usage. Although the improvement in one's own health status and life expectancy may simply reflect a third factor that "causes" both more schooling *and* better health, the absence of any obvious prior cause and the strength of the statistical relationship between schooling and these health-related outcomes suggests that one's own schooling may be the causal factor.<sup>13</sup>

Though some portion of the benefits of increased health status and life expectancy may be reflected in higher labor market earnings, it seems clear that nonmarket private gains from this relationship do exist (for example, consider the reduced pain and suffering, reduced anxiety in response to negative life events, reduced mortality, lower medical-care time and money expenditures). In addition, some of the benefits of one's own health improvements may be in the form of external benefits,

<sup>&</sup>lt;sup>11</sup> Two components of potentially important private nonmarket benefits are not included in the table. The first is the consumption value of schooling—the well-being that people experience from the process of attending school and the learning experience that is conveyed. The second is the consumer surplus associated with the benefits that are distinguished and that are valued by their implicit market price. (See the Appendix for a discussion of this source of well-being.) We were unable to identify empirical studies assessing the magnitude and value of gains from either of these sources.

<sup>&</sup>lt;sup>12</sup> Although economists are reluctant to claim the existence of a causal link, recent studies suggest that persons with more schooling are less likely to smoke, and among persons who do smoke, those with more schooling smoke less per day. An additional year of schooling reduces average daily cigarette consumption by 1.6 for men and 1.1 for women. People with more education are also less likely to be heavy drinkers and tend to engage in more exercise per week (about 17 minutes for each additional year of schooling) than are less-educated people (see Kenkel 1991).

<sup>&</sup>lt;sup>13</sup> A study using sibling data from Nicaragua finds evidence in both fixed and random effects models that the relationship between more schooling and better health is not due to unobserved or unmeasured factors, but is, in fact, causal (Behrman and Wolfe 1987).

ranging from the reduced spread of contagious disease to increased utility of relatives and friends whose well-being depends on one's own health.

An additional benefit accruing to the better-schooled individual comes in the form of increased knowledge and savvy regarding market transactions, referred to as "consumer efficiency" (category 8). Michael (1982) translates the finding that a person with an additional year of schooling is significantly more efficient as a consumer into dollars of additional income. Similarly, Benham and Benham (1975), analyzing the market for eyeglasses, find that persons with more schooling tended to pay less for glasses than those with less schooling; Morton, Zettelmeyer, and Silva-Risso (2001) report similar findings for the price paid for new cars. Rizzo and Zeckhauser (1992) find that the charge per unit of time that a physician spent with a patient was lower for better-educated individuals than for those with less education.

Categories 9 through 11 in Table 3 refer to the linkages between one's success in making choices involving the labor market, marriage, and family size and the level of schooling. In all of these cases more schooling is positively related to the quality of choices, perhaps through information gains through schooling that promote more efficient decisions. Part of this gain may be simply in the ability to accomplish better matches—in the labor market, for example—but another part may be in the reduction of time spent in the search.<sup>14</sup> Royalty (1998) provides evidence on another outcome associated with labor market efficiency-namely that for women, more schooling is associated with lower job turnover. Studies of assortative mating suggest that schooling is associated with "better" choices regarding marital partners (Becker, Landes, and Michael 1977) and with lower rates of divorce (Martin 2002). Better-educated people also tend to be more successful in securing desired family sizes; more schooling may enable one to gather information on how to avoid unwanted births and possibly also to reduce the probability of subfecundity. Evidence of this relationship also exists for developing countries.

### External (to the Household) and Public Goods Effects (Categories 12 through 17)

Beyond the gains to one's self and family are those seldom-noted and rarely evaluated external and public goods effects of one's education that

<sup>&</sup>lt;sup>14</sup> In addition to the individual, employers may also gain if more-schooled individuals yield a superior labor market match. Improved matching of employees to jobs reduces a variety of costs that are otherwise borne by employers, including training costs, recruiting costs, and the loss of productivity during employment transitions. Acemoglu and Angrist (1999) seem to include such effects in their effort to measure aspects of the gain from schooling beyond those included in the traditional return estimates.

accrue to others in society. There is evidence that the amount of time and money devoted to charity is positively associated with the amount of schooling one has, after controlling for income, the other primary determinant of donations (category 12). For example, one study found that college graduates volunteered nearly twice as many hours and donated 50 percent more of their income than high school graduates (see Hodgkinson and Weitzman 1988). The positive contribution of schooling to savings (category 13) may have a public-good aspect to the extent that the capital market is imperfect and aggregate savings are less than optimal. Similarly, increased education may lead to social cohesion and may enable one to better accommodate technological and social change (categories 14 and 15). Persons with more schooling may make moreinformed choices when voting and may participate more fully in their communities. Persons with more schooling may contribute to the common good in other ways. For example, there is evidence that schooling is positively related to being more trusting of others, to having an increased participation in community organizations (Helliwell and Putnam 1999), and to having a higher probability of nonviolent protests against government-sponsored repression (Hall, Rodeghier, and Useem 1986).

There is evidence that more schooling is associated with a lower probability of receiving transfer benefits, either disability-related benefits or welfare (category 16). Recent analyses have found that higher education of mothers reduces the probability that their daughters will, if eligible, elect to receive welfare benefits. Criminal activity in the community tends to be negatively related to the average level of educational attainment of members of the community (category 17).

The relationships listed in categories 3 through 17 represent potential effects of schooling that are not captured in traditional estimates of the private economic returns to education. We have characterized these as private nonmarket and external/public goods effects attributable to education. In all cases, research studies document the direction of the relationship, and in some cases its magnitude. To be sure, in some cases the strength of the evidence is less strong than one would desire. Among the most robust and substantial influences are the relationships between parents' schooling and the levels of health, schooling, and childbearing of their children. The linkages between one's own schooling and own health are also well documented. One is left with the impression that schooling has substantial benefits beyond those usually tabulated by measures of labor market productivity and fringe benefits.

# ON ESTIMATING THE VALUE OF NONMARKET IMPACTS OF EDUCATION

To translate these private nonmarket and external/public goods benefits into information relevant for public sector decisions on the allocation of resources to education, the value of each of the separate categories of effect, and of the entire bundle of these effects, must be estimated. In Haveman and Wolfe (1984), we developed a method to estimate the marginal value of schooling attributable to the private nonmarket components of these effects (categories 3 through 11). This method is based on a traditional household production function relationship that relates the contributions of schooling and market inputs in producing nonmarket outcomes. Consumers, acting as firms, efficiently combine inputs, including schooling services, so as to yield a consumption frontier for goods or services that enter their utility function. These consumers maximize utility subject to this consumption frontier. Studies that establish a reliable relationship between education plus some other input that carries a market price, and a nonmarket outcome-such as health, consumer efficiency, educational attainment of children, and so forth-can be used with this method to generate estimates of the marginal value of schooling.

To implement this approach, each of these studies must have a coefficient estimate relating schooling to the outcome of interest, as well as controls for other additional variables likely to be associated with that outcome. In addition, each study must include another input to "production" of the nonmarket outcome of interest that has a market value that is not subject to market imperfections. In addition, when this input is used in the "production" of the outcome of interest, its use must be exclusive-that is, the amount of the input used in producing the outcome of interest is "used up" in the production of the output. Examples of inputs with such a market value might include physician visits, spending on police in the community, private music lessons, and so forth. When such inputs are not available, income may be used under the assumption that income will be spent on the output only until the marginal product per dollar spent is equal to that of other inputs including schooling. The coefficient on income then represents the marginal product of income spent on the outcome under study.

The following simplified model illustrates this approach, using a single nonmarket good. The model makes the standard economic assumption that individuals or households efficiently combine schooling with other market inputs to produce the nonmarket outcome. A wellknown result in economics is that efficient producers will equate the ratio of the marginal product to input price, across all inputs. This relationship also holds in the production of the nonmarket outcome, with schooling and at least one other market input. That is,

$$\frac{MP_{SCH}}{P_{SCH}} = \frac{MP_X}{P_X},\tag{1}$$

where  $MP_{SCH}$  is the marginal product of schooling in producing the nonmarket outcome,  $MP_X$  is the marginal product of any input X with market price  $P_X$ , and  $P_{SCH}$  is the implicit price or willingness to pay for additional schooling in producing the nonmarket outcome. A little rearranging yields the following formula for computing the implicit price or willingness to pay for additional schooling in producing a nonmarket outcome:

$$P_{SCH} = \frac{MP_{SCH}}{MP_{X}} \times P_{X.}$$
<sup>(2)</sup>

This equation for the implicit value of additional schooling in producing a particular private nonmarket output is intuitively appealing. If the marginal products of schooling and the other input are equal, the implicit willingness to pay for the effect of schooling on this outcome will be equal to the price of the other input. If the marginal product of schooling is double that of the other input, the implicit value of schooling is twice the unit price of the other input.<sup>15</sup>

Implementing this method involves estimating the productive relationship ( $MP_{SCH}$ ) between schooling and each private nonmarket outcome. It also requires estimating the productive relationship ( $MP_x$ ) between each outcome measure and another input. The latter input should be one that is competitively marketed. Once these marginal productivities are estimated, they can be combined with the private cost of the privately purchased input in order to estimate the implicit willingness to pay for additional schooling for each outcome, using the formula given in equation (2). The implicit value for each individual outcome can then be summed to produce the total incremental value of additional schooling.

This approach, it should be noted, requires that several conditions hold if the estimates are to be reliable. A brief listing of them here will make clear the tentative nature of estimates obtained from applying this method. First, consumers must not be constrained in their choice of homogeneous schooling services and market inputs in producing the private nonmarket good or service. Second, the value of the market input must reflect the operation of a smoothly functioning competitive market; only if this holds will the value imputed into marginal units of schooling reflect a willingness to pay as it would be revealed in a market. Third, it

<sup>&</sup>lt;sup>15</sup> Extension of the simple model presented here to the production of multiple nonmarket and market outcomes, such as wage income, is straightforward (see Haveman and Wolfe 1984). The total willingness to pay for additional schooling across all nonmarket and market outcomes is the sum of the implicit willingness to pay for each individual outcome. Our fully developed model accounts for the nonexclusivity (nondivisibility) of schooling in producing multiple outcomes.

must be assumed that the composition of other inputs in the production process does not change with changes in schooling. That is, the gains attributed to schooling must reflect the direct increase in the productivity of labor and not the potential effects of schooling in improving the efficiency with which resources are combined in producing the output, or any changes in the amount of time more-schooled individuals spend in a given activity.<sup>16</sup> Finally, the empirical studies on which this model rests must provide estimates of linkages both of schooling and of the market input to the nonmarket output that are not biased and inconsistent because of unobserved characteristics. This is a strong assumption, given the single-equation regression framework that underlies most of the studies that we cite. Hence, this method must be viewed as yielding a first-cut approximation of the private nonmarket values and as a guide for further research.<sup>17</sup>

While recognizing these assumptions, we nevertheless use this approach to generate a few first-cut estimates of the value of nonmarket impacts; these are shown in Table 4. We convert a small number of impacts into the marginal relationship, or further into a willingness-to-pay estimate. We base our results on coefficients obtained from the studies listed in the third column of the table. Estimates of the implicit value of the private nonmarket effects of schooling are provided for the cognitive development of children (category 4),<sup>18</sup> consumption

<sup>&</sup>lt;sup>16</sup> Welch (1970) first discusses this important distinction. A discussant of this paper argues that this condition is not likely to hold, resulting in overestimates of the value of the private nonmarket outputs of schooling. Moreover, if it does not hold, the value of the additional resources used in the reorganized production process must be accounted for in the calculation of the net value of incremental schooling. For further development of this point, see Rosenzweig and Schultz (1982).

<sup>&</sup>lt;sup>17</sup> Paul Schultz, in his discussion of this paper, states: "I look forward to a new generation of empirical research into the role of education in household production, from which more adequate and less biased evaluations of the nonmarket returns to education can be derived using the conceptual logic [of this method]." Such research might be similar to the approaches used in trying to better identify the market private returns to schooling discussed above, including the use of data on siblings and the application of instrumental variable techniques.

<sup>&</sup>lt;sup>18</sup> For example, a recent study by Ermisch and Francesconi (2000) and the special tabulations of Ermisch (1999) provide estimates of the impact of mother's education and household income on the level of schooling achieved by children (category 4) in the United Kingdom, using data drawn from the British Household Panel Study. The coefficient estimate for household income (the input with market values) is 0.098 (t-statistic = 1.668) for girls, indicating that, at the margin, an additional dollar of household income increases the expected level of schooling. The mother's education is represented by dummy variables for six levels of schooling ranging from less than O level to first and higher (with no qualification as the omitted category) in an ordered logit estimation. The simulation of the effects of a mother's education and family income (at the youngest age that they observe it, mainly around age 16) on the distribution of a daughter's qualifications is 0.218 for a mother's vocational degree on the probability that the daughter will have a vocational degree, and 0.255 for a mother's first or higher degree on the probability of the child having a vocational degree, while the relationship of family income to a vocational degree is 0.187.

Table 4

Estimates of the Annual Value	(Willingness to Pay)	or Impact of Additional	Schooling

Outcome	Value or Impact	Source of Coefficients
Cognitive Development of Children	\$350 in family income for high school diploma (vs. no diploma) and \$440 for some college (vs. high school diploma)	Angrist and Lavy (1996) <sup>a</sup>
	\$860-\$5,175 per year in future family income for an additional year of schooling	*Murnane (1981) <sup>b</sup> ; *Edwards and Grossman (1979) <sup>c</sup>
	£1166–£1727 in family income for mother's educational attainment of vocational/first and higher degrees	Ermisch (1999)
	\$4,008 in permanent family income for an increase in 4.8 years of grandfather's schooling; \$2,692 in permanent family income for an increase in 3.6 years of grandmother's schooling	Blau (1999)
Consumption Efficiency	<ul> <li>\$290 in household income for an additional year of schooling; save approximately</li> <li>\$5.50 per pair of eyeglasses for an additional year of schooling</li> </ul>	*Michael (1975); *Benham and Benham (1975) <sup>d</sup>
Own Health	\$8,950 in increased net family assets for an additional year of schooling	*Lee (1982)
	1.6 (1.1) fewer cigarettes smoked per day by men (women) for an additional year of schooling; and 34 more minutes of exercise per two weeks	Kenkel (1991) <sup>e</sup>
	1.85 (1.25) (1.37) greater relative risk of death from heart disease for males aged 45–64 (males aged 65–74) (females aged 65–74) with 8–11 years of schooling compared with those with 12 or more years of schooling	Feldman et al. (1989) <sup>f</sup>
Reduction in Criminal Activity	\$170 reduction in per capita expenditure on police for an additional mean year of schooling in community	*Ehrlich (1975)
Volunteer Hours	\$51 for males per year; \$30 for females per year	Freeman (1997)

Source: Estimates indicated by an asterisk (\*) are taken from Haveman and Wolfe (1984), Table 2. All other values and impacts are based on coefficients in studies listed in the third column. All values are in 1996 dollars except as noted.

<sup>a</sup> Based on National Longitudinal Survey of Youth (Table 8, column 4 estimates).

<sup>b</sup> Based on measurement of cognitive development on Iowa Test of Basic Skills using children in grades three through six whose families participated in the Negative Income Tax experiment in Gary, Indiana. For conversion see Haveman and Wolfe (1984).

 $^{\rm c}$  Based on data from cycle II of the Health Examination Survey using the mean of the estimated value of the mother's and father's education.

 $^{\rm d}$  Based on 1970 Health Interview Survey (HIS); n = 10,000, of which 1,625 obtained eyeglasses in 1970.

 $^{\rm e}$  Based on 1985 Supplement to the HIS on Health Promotion and Disease Prevention; n=14,177 males and 19,453 females.

<sup>f</sup> Based on 62,405 persons in Matched Records Study, whites only (see Feldman et al. 1989).

efficiency (category 8), own health (category 7), reduction in criminal activity (category 17), and charitable giving (volunteer hours) (category 12).

The estimates of these privately captured, though nonmarket, benefits shown in Table 4 are substantial. The willingness to pay for the cognitive development gains of one's children attributable to an additional year of one's own schooling vary substantially, but it is not unreasonable to impute an average annual family gain of at least \$500. Improvements in the efficiency of consumer choices attributable to another year of schooling would seem to convey an average of at least \$300 per year in benefits. The value of the improvement in one's own health from additional schooling seems substantial—a onetime payment of several thousand dollars for an additional year of schooling. Somewhat smaller, though not trivial, annual gains are also attributed to the reduction in criminal activity associated with additional schooling in a community and the willingness-to-pay value of the additional volunteer activity associated with a year of incremental schooling.

It is not unreasonable to suggest that, when the social gains from all of the categories of private nonmarket and external/public goods identified in Table 3 are taken into account, their sum could equal estimates of the annual earnings impacts of an additional year of schooling<sup>19</sup>

$$P_{SCH} = \frac{MP_{SCH}}{MP_X} \times P_X = \frac{.218}{.187} \times \pounds 1,000 = \pounds 1,166.$$

This translates into a pound value of £1,166. The pound value of a mother's first or higher degree on the probability that the daughter will have a vocational degree, in terms of annual family income, is £1,364. Similarly, the simulated value of a mother's vocational degree on the daughter's first or higher degree is 0.107, and the simulated value for a mother's having a first or higher degree on her daughter's achieving the same degree is 0.152. In this case, family income has a simulated relationship of 0.088, providing economic estimates of a mother's additional schooling of £1,216 and £1,727, respectively (both levels of schooling for the mother are statistically significant at the 1 percent level). In this estimate as well as others in Table 4, we use earnings or income as the variable to infer the willingness to pay for incremental schooling. It should be noted that these variables may be endogenous to the labor supply choices; wage rates would be a preferable variable on which to base our estimate, but they are seldom reported. Moreover, in using family income or assets, as in increase in education to the family and not only to the person whose education is being varied. (We thank Bruce Chapman for pointing out this implication to us.)

<sup>19</sup> These annual economic gains in the form of increased earnings attributable to an additional year of schooling are on the order of \$2,000 to \$4,000 per year, depending on the study (see, for example, Figures 1 and 2).

Using the formula of equation (2), we derive the marginal value of a mother's vocational degree on the probability that the daughter will have a vocational degree, in terms of annual family income, as follows:

captured in the traditional returns-to-schooling studies.<sup>20</sup> If that is the case, the full social rate of return to an additional year of schooling could be twice the private economic rates of return to education—ranging from about 7 to 9 percent—estimated in the traditional studies.

#### POLICY IMPLICATIONS AND CONCLUSIONS

The conclusion that the full social gains from additional schooling exceed—perhaps substantially—the 7 to 9 percent private rate found in the returns-to-schooling literature has important implications for public policy. Because of these private nonmarket and external/public goods effects, the answers to two central policy questions should be fundamentally assessed. These two questions are:

- What volume of the nation's resources should be allocated to the production of schooling services?
- Who should be paying for these schooling services?

If the nonmarket private and external/public goods effects of education are equal in value to the private market returns, the full rate of return could be as high as 14 to 18 percent. Because few other public or private investments seem able to claim returns of this magnitude, a reallocation of resources from other uses to the education sector may be in order.

This leaves open the question of who should bear the cost of the efficient level of schooling services. There are two primary candidates the private citizens who receive these services (and their families) and the public sector. While our results suggest that the full value of the private nonmarket and external/public goods effects of education may be substantial and, hence, should be reflected in resource allocation decisions, they say little about the balance between the total privately captured benefits and the spillover/public goods components of the full social benefits of education.

One might conclude from our calculations that the private nonmarket gains from education are substantial, leading to the judgment that a greater share of the full social benefits of schooling is captured by students and their families than is suggested by the traditional economic returns estimates. If that were so, the case for increases in tuition and fees as an efficient means of financing schooling, especially at the higher educational level, would be strengthened. Of course, an increase in the

<sup>&</sup>lt;sup>20</sup> To our knowledge, very few estimates of the social returns to schooling exist. One of the few attempts uses an instrumental variable approach to attempt to capture labor market productivity beyond that captured in private worker-based rates of return. This approach is akin to efforts to identify the social returns to schooling reflected in the endogenous growth literature. See Acemoglu and Angrist (1999).

price charged for market purchases of schooling services implies little regarding the nature and magnitude of targeted student aid designed to increase educational opportunities for those students lacking the resources necessary to pay for these services (and constrained from borrowing to pay for them).

#### References

- Acemoglu, Daron and Joshua Angrist. 1999. "How Large Are the Social Returns to Education? Evidence from Compulsory Schooling Laws." NBER Working Paper No. 7444 (December).
- Allen, Steven G. 2001. "Technology and the Wage Structure." *Journal of Labor Economics* 19 (2): 440–83.
- An, Chong Bum, Robert H. Haveman, and Barbara L. Wolfe. 1993. "Teen Out-of-Wedlock Births and Welfare Receipt: The Role of Childhood Events and Economic Circumstances." *Review of Economics and Statistics* 75 (2): 195–208.
- Angrist, Joshua D. and Victor Lavy. 1996. "The Effect of Teen Childbearing and Single Parenthood on Childhood Disabilities and Progress in School." NBER Working Paper No. 5807 (October).
- Antel, John. 1988. "Mother's Welfare Dependency Effects on Daughter's Early Fertility and Fertility Out of Wedlock." University of Houston, working paper.
- Ashenfelter, Orley and Cecilia Rouse. 1998. "Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins." *Quarterly Journal of Economics* 113 (1): 253–84.
- Ashenfelter, Orley, Colm Harmon, and Hessel Oosterbeek. 2000. "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." NBER Working Paper No. 7457 (January).
- Auster, Richard, Irving Leveson, and Deborah Sarachek. 1969. "The Production of Health: An Exploratory Study." *Journal of Human Resources* 4 (4): 411–36.
- Autor, David, Lawrence Katz, and Alan Krueger. 1998. "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics* 113 (4): 1169–1213.
- Axinn, William, Greg Duncan, and Arland Thornton. 1997. "The Effects of Parents' Income, Wealth and Attitudes in Children's Completed Schooling and Self-Esteem." In *The Consequences of Growing Up Poor*, edited by G. Duncan and J. Brooks-Gunn. New York: Russell Sage.
- Barro, Robert J. 1997. Determinants of Economic Growth: A Cross-Country Empirical Study. Cambridge, MA: MIT Press.
- Barro, Robert J. and Jong-Wha Lee. Forthcoming. "International Data on Educational Attainment: Updates and Implications." *Oxford Economic Papers.*
- Bartel, Ann and Nachum Sicherman. 1999. "Technological Change and Wages: An Interindustry Analysis." Journal of Political Economy 107 (2): 285–325.
- Becker, Gary S. 1964. Human Capital: A Theoretical and Empirical Analysis. New York: Columbia University Press (for NBER).
- Becker, Gary S., Elizabeth M. Landes, and Robert T. Michael. 1977. "An Economic Analysis of Marital Instability." *Journal of Political Economy* 85 (6): 1141–88.
- Behrman, Jere R. and Nevzer Stacey, eds. 1997. *The Social Benefits of Education*. Ann Arbor: University of Michigan Press.
- Behrman, Jere R. and Barbara L. Wolfe. 1987. "How Does Mother's Schooling Affect Family Health, Nutrition Medical Care Usage, and Household Sanitation?" *Journal of Econometrics* 36 (1/2): 195–204.
- Behrman, Jere R., Andrew D. Foster, Mark R. Rosenzweig, and Prem Vashishtha. 1999. "Women's Schooling, Home Teaching, and Economic Growth." *Journal of Political Economy* 107 (4): 682–714.
- Benham, Lee. 1974. "Benefits of Women's Education within Marriage." In Economics of the

*Family: Marriage, Children, and Human Capital,* edited by T. W. Schultz. Chicago: University of Chicago Press (for NBER).

- Benham, Lee and Alexandria Benham. 1975. "Regulating through the Professions: A Perspective on Information Control." *Journal of Law and Economics* 18 (2): 421–47.
- Berger, Mark and J. Paul Leigh. 1989. "Schooling, Self-Selection, and Health." Journal of Human Resources 24 (3): 433–55.
- Blau, David. 1999. "Effect of Income on Child Development." Review of Economics and Statistics 81 (2): 261–76.
- Bound, John and George Johnson. 1992. "Changes in the Structure of Wages during the 1980s: An Evaluation of Alternative Explanations." *American Economic Review* 82 (3): 371–92.
- Bowles, Samuel, Herbert Gintis, and Melissa Osborne. 2001. "The Determinants of Earnings: A Behavioral Approach." *Journal of Economic Literature* 34 (4): 1137–76.
- Campbell, Angus, Phillip E. Converse, Warren E. Miller, and Donald E. Stokes. 1976. *The American Voter*. Chicago: University of Chicago Press.
- Card, David. 2001. "Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69 (5): 1127–61.
- Clark, Rebecca. 1992. "Neighborhood Effects on Dropping Out among Teenage Boys." Urban Institute Working Paper PSC-DSC-UI-13. Washington, DC: Urban Institute.
- Comer, James P. 1988. "Educating Poor Minority Children." Scientific American 29 (5): 42-8.
- Conlisk, John. 1971. "A Bit of Evidence on the Income-Education-Ability Interrelation." Journal of Human Resources 6 (3): 358–62.
- Corman, Hope. 1986. "The Demand for Education for Home Production." *Economic Inquiry* 24 (2): 213–30.
- Crimmins, Eileen M. and Yasuhiko Saito. 2001. "Trends in Healthy Life Expectancy in the United States, 1970–1990: Gender, Racial, and Education Differences." Social Science and Medicine 52 (11): 1629–41.
- DaVanzo, Julie. 1983. "Repeat Migration in the United States: Who Moves Back and Who Moves On?" *Review of Economics and Statistics* 65 (4): 552–9.
- Dawson, Deborah. 1991. "Family Structure and Children's Health and Well-Being: Data from the 1988 National Health Interview Survey on Children's Health." Journal of Marriage and the Family 53 (3): 373–84.
- Duncan, Greg J. 1976. "Earnings Functions and Nonpecuniary Benefits." Journal of Human Resources 11 (3): 464–83.

——. 1994. "Families and Neighbors as Sources of Disadvantage in the Schooling Decisions of White and Black Adolescents." *American Journal of Education* 103 (1): 20–53.

- Duniform, Rachel, Greg J. Duncan, and Jeanne Brooks-Gunn. 2000. "As Ye Clean, so Shall Ye Glean: Some Impacts of 'Non-Cognitive' Characteristics within and across Generations." Paper presented at annual meeting of American Economic Association, January 2001.
- Dye, Richard F. 1980. "Contributions to Volunteer Time: Some Evidence on Income Tax Effect." National Tax Journal 33 (1): 89–93.
- Easterlin, Richard A. 1968. Population, Labor Force, and Long Swings in Economic Growth: The American Experience. New York: NBER.
- Edgmand, Michael R., Ronald L. Moomaw, and Kent W. Olson. 1994. "College Education: What Is It Worth?" In *Economics and Contemporary Studies*, by M. R. Edgmand, R. L. Moomaw, and K. W. Olson. Chicago: Dryden Press.
- Edwards, Linda N. and Michael Grossman. 1979. "The Relationship between Children's Health and Intellectual Development." In *Health: What Is It Worth?* edited by S. Mushkin. Elmsford, NY: Pergamon Press.
- Ehrlich, Isaac. 1975. "On the Relation between Education and Crime." In *Education, Income, and Human Behavior*, edited by F. T. Juster. New York: McGraw-Hill.
- Ermisch, John. 1999. Unpublished data tables prepared for the authors that extend Ermisch and Francesconi (1997).
- Ermisch, John and Marco Francesconi. 1997. "Family Matters." CEPR Discussion Paper No. 1591, Centre for Economic Policy Research, London.
- 2000. "Educational Choice, Families and Young People's Earnings." Journal of Human Resources 35 (1): 143–76.
- Feldman, Jacob, Diane Makuc, Joel Kleinman, and Joan Cornoni-Huntly. 1989. "National

Trends in Educational Differences in Mortality." American Journal of Epidemiology 129 (5): 919–33.

- Foster, Andrew D. and Mark R. Rosenzweig. 1996. "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution." *American Economic Review* 86 (4): 931–53.
- Freeman, Richard B. 1981. "The Effect of Unionism on Fringe Benefits." Industrial & Labor Relations Review 34 (4): 489–509.
  - —. 1995. "The Labor Market." In *Crime*, edited by J. Q. Wilson and J. Petersilia. San Francisco: Institute for Contemporary Studies.
- ——. 1997. "Working for Nothing: The Supply of Volunteer Labor." Journal of Labor Economics 15 (1): S140-66.
- Ginther, Donna, Robert Haveman, and Barbara Wolfe. 2000. "Neighborhood Attributes as Determinants of Children's Outcomes: How Robust Are the Relationships?" *Journal of Human Resources* 35 (4): 603–42.
- Gintis, Herbert. 1971. "Education, Technology, and the Characteristics of Worker Productivity." American Economic Review 61 (2): 266–79.
- Glewwe, Paul. 1999. "Why Does Mother's Schooling Raise Child Health in Developing Countries? Evidence from Morocco." *Journal of Human Resources* 34 (1): 124–59.
- Greenwood, Daphne T. 1997. "New Developments in the Intergenerational Impact of Education." International Journal of Educational Research 27 (6): 503–11.
- Greenwood, Michael J. 1975. "Research on Internal Migration in the United States: A Survey." *Journal of Economic Literature* 13 (2): 397–433.
- Griliches, Zvi. 1977. "Estimating the Returns to Schooling: Some Econometric Problems." Econometrica 45 (1): 1–45.
- Grossman, Michael. 1975. "The Correlation between Health and Schooling." In *Household Production and Consumption*, edited by N. E. Terleckyj. New York: NBER.
- Grossman, Michael and Steven Jacobowitz. 1981. "Variations in Infant Mortality Rates among Counties in the United States: The Roles of Public Policies and Programs." *Demography* 18 (4): 695–713.
- Grossman, Michael and Theodore Joyce. 1989. "Socio-Economic Status and Health: A Personal Research Perspective." In *Pathways to Health: The Role of Social Factors*, edited by J. Bunker, D. Genby, and B. Kehrer. Menlo Park, CA: Kaiser Foundation.
- Hall, Robert, Mark Rodeghier, and Bert Useem. 1986. "Effects of Education on Attitude to Protest." *American Sociological Review* 51 (4): 564–73.
- Hansen, W. Lee. 1963. "Total and Private Rates of Return to Investment in Schooling." Journal of Political Economy 71 (2): 128-40.
- Hauser, Seth. 2000. "Education, Ability and Civic Engagement in the Contemporary United States." Social Science Research 29 (4): 556–82.
- Haveman, Robert H. and Barbara L. Wolfe. 1984. "Schooling and Economic Well-Being: The Role of Nonmarket Effects." *Journal of Human Resources* 19 (4): 378–407.
- ——. 1994. Succeeding Generations: On the Effects of Investments in Children. New York: Russell Sage.
- ——. 2001. "Accounting for the Social and Non-Market Benefits of Education." In *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being*, edited by J. Helliwell. OECD/Human Resources Development Canada. Vancouver: University of British Columbia Press.
- Haveman, Robert H., Barbara L. Wolfe, and James Spaulding. 1991. "Childhood Events and Circumstances Influencing High School Completion." *Demography* 28 (1): 133–57.
- Haveman, Robert H., Barbara L. Wolfe, and Kathyrn Wilson. 2001. "The Role of Economic Incentives in Teenage Nonmarital Childbearing Choices." *Journal of Public Economics* 81 (3): 473–511.
- Hayward, Mark D., William Grady, and John O. Billy. 1992. "The Influence of Socioeconomic Status on Adolescent Pregnancy." Social Science Quarterly 73 (4): 750–72.
- Helliwell, John F., and Robert D. Putnam. 1999. "Education and Social Capital." NBER Working Paper No. 7121 (May).
- Herzog, A. Regula, Melissa Franks, Hazel Markus, and Diane Holmberg. 1998. "Activities and Well-Being in Older Age: Effects of Self-Concept and Educational Attainment." *Psychology and Aging* 13 (2): 179–85.
- Hodgkinson, Virginia A. and Murray S. Weitzman. 1988. Giving and Volunteering in the

United States: Findings from a National Survey, 1988 Edition. Washington, DC: Independent Sector.

- Kemna, Harrie J. M. I. 1987. "Working Conditions and the Relationship between Schooling and Health." Journal of Health Economics 6 (3): 189-210. Kenkel, Donald. 1991. "Health Behavior, Health Knowledge, and Schooling." Journal of
- Political Economy 99 (2): 287-305.
- Kiefer, Nicholas M. 1985. "Evidence on the Role of Education in Labor Turnover." Journal of Human Resources 20 (3): 445-52.
- King, Elizabeth and M. Anne Hill, eds. 1993. Women's Education in Developing Countries. Baltimore: Johns Hopkins University Press.
- Lam, David and Suzanne Duryea. 1999. "Effects of Schooling on Fertility, Labor Supply, and Investments in Children, with Evidence from Brazil." Journal of Human Resources 34 (1): 160-92.
- Lee, Lung Fei. 1982. "Health and Wage: A Simultaneous Equation Model with Multiple Discrete Indicators." International Economic Review 23 (1): 199-222.
- Leigh, J. Paul. 1981. "Hazardous Occupations, Illness, and Schooling." Economics of Education Review 1 (3): 381–88.
- -. 1983. "Direct and Indirect Effects of Education on Health." Social Science and Medicine 17 (4): 227-34.
- Light, Audrey. 2001. "In-School Work Experience and the Returns to Schooling." Journal of Labor Economics 19 (1): 65–93.
- Lochner, Lance and Enrico Moretti. 2001. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." NBER Working Paper No. 8605 (November).
- Lucas, Robert E. B. 1977. "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling." American Economic Review 67 (4): 549-58.
- Mansfield, Edwin. 1982. "Education, R and D, and Productivity Growth." National Institute of Education Special Report, Washington, DC.
- Martin, Steve. 2002. "Marital Dissolutions Involving Young Children: Trends by Education and Race since 1970." Paper presented at the Institute for Research on Poverty Low Income Workshop, Madison, WI.
- McMahon, Walter. 2000. Education and Development: Measuring the Social Benefits. New York: Oxford University Press.
- Metcalf, David. 1973. "Pay Dispersion, Information, and Returns to Search in a Professional Labour Market." Review of Economic Studies 40 (4): 491-505.
- Michael, Robert T. 1972. The Effect of Education on Efficiency in Consumption. New York: Columbia University Press (for NBER).
  - . 1975. "Education and Consumption." In Education, Income, and Human Behavior, edited by F. T. Juster. New York: McGraw-Hill.
  - -. 1982. "Measuring Non-Monetary Benefits of Education: A Survey." In Financing Education: Overcoming Inefficiency and Inequity, edited by W. McMahon and T. Geske. Urbana, IL: University of Illinois Press.
- Michael, Robert T. and Robert J. Willis. 1976. "Contraception and Fertility: Household Production under Uncertainty." In Household Production and Consumption, edited by N. E. Terleckyj. Studies in Income and Wealth No. 40. New York: NBER.
- Mincer, Jacob. 1962. "On-the-Job Training: Costs, Returns, and Some Implications." Journal of Political Economy 70 (5) part 2: 50-79.
- Morton, Fiona Scott, Florian Zettelmeyer, and Jorge Silva-Risso. 2001. "Consumer Information and Price Discrimination: Does the Internet Affect the Pricing of New Cars to Women and Minorities?" NBER Working Paper No. 8668 (December).
- Mueller, Marnie W. 1978. "An Economic Theory of Volunteer Work." Department of Economics, Wesleyan University, unpublished paper, Middletown, CT.
- Muller, Andreas. 2002. "Education, Income Inequality, and Mortality: A Multiple Regression Analysis." British Medical Journal 324 (7328): 23-25.
- Murnane, Richard J. 1981. "New Evidence on the Relationship between Mother's Education and Children's Cognitive Skills." Economics of Education Review 1 (2): 245-52.
- Nelson, Richard R. 1973. "Recent Exercises in Growth Accounting: New Understanding or Dead End?" American Economic Review 63 (3): 462-8.
- Neuman, Shoshana and Adrian Ziderman. 1990. "Does A Woman's Education Affect Her

Husband's Earnings? Results for Israel in a Dual Labor Market." World Bank Policy Research Working Paper 464.

Organization for Economic Cooperation and Development. 1996. *Education at a Glance*. Centre for Educational Research and Innovation. Paris: OECD.

—. 1997. Annual National Accounts. Vol. 1. Paris: OECD.

———. 1998. *Education at a Glance*. Centre for Educational Research and Innovation. Paris: OECD.

——. 1999. *Education at a Glance*. Centre for Educational Research and Innovation. Paris: OECD.

———. 2000. *Education at a Glance*. Centre for Educational Research and Innovation. Paris: OECD.

Pascarella, Ernest T. and Patrick T. Terenzini. 1991. *How College Affects Students*. San Francisco: Jossey-Bass Publishers.

Pauly, Mark. 1980. Doctors and Their Workshops: Economic Models of Physician Behavior. Chicago: University of Chicago Press (for NBER).

Reynolds, Arthur. 2000. Success in Early Intervention: The Chicago Child-Parent Centers. Lincoln, NE: University of Nebraska Press.

Ribar, David C. 1993. "A Multinomial Logit Analysis of Teenage Fertility and High School Completion." *Economics of Education Review* 12 (2): 153–64.

- Rizzo, John A. and Richard Zeckhauser. 1992. "Advertising and the Price, Quantity, and Quality of Primary Care Physician Services." *Journal of Human Resources* 27 (3): 381–421.
- Robins, Lee N. 1984. "Lifetime Prevalence of Specific Psychiatric Disorders in Three Sites." Archives of General Psychiatry 41 (10): 949–58.
- Rosenzweig, Mark R. and T. Paul Schultz. 1982. "Education and the Household Production of Child Health." Social Statistics Section Proceedings of the American Statistical Association. Washington, DC.

—. 1983. "Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs and the Effects on Birthweight." *Journal of Political Economy* 91 (5): 723–46.

—. 1989. "Schooling, Information, and Nonmarket Productivity: Contraceptive Use and Its Effectiveness." *International Economic Review* 30 (2): 457–77.

Ross, Catherine E. and Chia-Ling Wu. 1988. "Education, Age and the Cumulative Advantage in Health." *Journal of Health and Social Behavior* 37 (1): 104–20.

——. 1995. "The Links between Education and Health." American Sociological Review 60 (5): 719–45.

- Royalty, Anne Beeson. 1998. "Job-to-Job and Job-to-Nonemployment Turnover by Gender and Educational Level." Journal of Labor Economics 16 (2): 392–443.
- Rudd, Nancy M., Patrick C. McKenry, and Myungkyun Nah. 1990. "Welfare Receipt among Black and White Adolescent Mothers: A Longitudinal Perspective." *Journal of Family Issues* 11 (3): 334–52.

Ryder, Norman B., and Charles F. Westoff. 1971. *Reproduction in the United States*, 1965. Princeton, NJ: Princeton University Press.

- Sandefur, Gary and Sara McLanahan. 1990. "Family Background, Race and Ethnicity, and Early Family Formation." Institute for Research on Poverty, Discussion Paper No. 911-90, University of Wisconsin–Madison.
- Sandefur, Gary D., Sara McLanahan, and Roger A. Wojtkiewicz. 1989. "Race and Ethnicity, Family Structure, and High School Graduation." Discussion Paper 893-89, Institute for Research on Poverty, University of Wisconsin–Madison.
- Sander, William. 1995. "Schooling and Quitting Smoking." Review of Economics and Statistics 77 (1): 191–99.
- Schieman, Scott. 2000. "Education and the Activation, Course, and Management of Anger." Journal of Health and Social Behavior 41 (1): 20–39.
- Schultz, T. Paul. 2002. "Discussion: Social and Nonmarket Benefits from Education in an Advanced Economy," this volume.
- Schultz, Theodore W. 1961. "Investment in Human Capital." American Economic Review 51 (5): 1–17.
- Shakotko, Robert, Linda Edwards, and Michael Grossman. 1981. "An Exploration of the Dynamic Relationship between Health and Cognitive Development in Adolescence." In

*Contributions to Economic Analysis: Health, Economics, and Health Economics,* edited by J. van der Gaag and M. Perlman. Amsterdam: North-Holland.

- Sherman, Lawrence W., Denise Gottfredson, Doris L. MacKenzie, John Eck, Peter Reuters, and Shawn Bushway. 1998. "Preventing Crime: What Works, What Doesn't, What's Promising." NCJ Publication No. 165366. Washington, DC: National Institute of Justice.
- Smeeding, Timothy. 1983. "The Size Distribution of Wage and Nonwage Compensation: Employer Cost versus Employee Value." In *The Measurement of Labor Cost*, edited by J. Triplett. Chicago: University of Chicago Press.
- Smith, Judith, Jeanne Brooks-Gunn, and Pamela Klebanov. 1997. "Consequences of Living in Poverty for Young Children's Cognitive and Verbal Ability and Early School Achievement." In Consequences of Growing Up Poor, edited by G. Duncan and J. Brooks-Gunn. New York: Russell Sage.
- Solomon, Lewis C. 1975. "The Relation between Schooling and Savings Behavior: An Example of the Indirect Effects of Education." In *Education, Income, and Human Behavior*, edited by F. T. Juster. New York: McGraw-Hill.
- South, Scott and Eric Baumer. 2000. "Deciphering Community and Race Effects on Adolescent Premarital Childbearing." Social Forces 78 (4): 1379–1408.
- Strauss, John. 1990. "Households, Communities, and Preschool Children's Nutrition Outcomes: Evidence from Rural Côte d'Ivoire." *Economic Development and Cultural Change* 38 (2): 231–62.
- Strauss, John, Paul J. Gertler, Omar Rahman, and Kristin Fox. 1993. "Gender and Life-Cycle Differentials in the Patterns and Determinants of Adult Health." *Journal of Human Resources* 28 (4): 791–837.
- Thoits, Peggy A. 1984. "Explaining Distributions of Psychological Variability: Lack of Social Support in the Face of Life Stress." *Social Forces* 63 (2): 453–81.
- Thomas, Duncan, John Strauss, and Maria-Helena Henriques. 1991. "How Does Mother's Education Affect Child Height?" *Journal of Human Resources* 26 (2): 183–211.
- U.S. Department of Education. National Center for Education Statistics. 2001. *Digest of Education Statistics*, 2000. NCES 2001-034, by T. D. Snyder and C. M. Hoffman. Washington, DC.
- Vanness, David, and Barbara Wolfe. 2002. "Government Mandates and Employer-Sponsored Health Insurance: Who Is Still Not Covered?" International Journal of Health Care Finance and Economics 2 (2): 99–135.
- Wachtel, Paul. 1975. "The Effect of School Quality on Achievement, Attainment Levels, and Lifetime Earnings." Explorations in Economic Research 2 (4): 502–36.
- Welch, Finis. 1970. "Education in Production." Journal of Political Economy 78 (1): 35-59.
- Wolfe, Barbara L. and Jere R. Behrman. 1982. "Determinants of Child Mortality, Health, and Nutrition in a Developing Country." Journal of Development Economics 11 (2): 163–94.
- Wolfe, Barbara L. and Samuel Zuvekas. 1997. "Nonmarket Outcomes of Schooling." International Journal of Educational Research 27 (6): 491–502.
- Wolfinger, Raymond E., and Steven A. Rosenstone. 1980. *Who Votes*? New Haven, CT: Yale University Press.
- Wong, Yue-chim. 1986. "Entrepreneurship, Marriage and Earnings." Review of Economics and Statistics 68 (4): 693–99.
- Wozniak, Gregory D. 1987. "Human Capital, Information, and the Early Adoption of New Technology." *Journal of Human Resources* 22 (1): 101–12.
- Yamada, Tadashi, Tetsuji Yamada, and John M. Kang. 1991. "Crime Rates versus Labor Market Conditions: Theory and Time-Related Evidence." NBER Working Paper No. 3801 (August).

## APPENDIX: THE CONCEPT OF HUMAN CAPITAL AND ITS VALUE: A FRAMEWORK

In this appendix, we view education and schooling as a form of human capital. We present a framework for thinking about the value to society of both a stock of human capital—say, a person with a given level of skill and education—and the social value of an investment in human capital. This framework is comprehensive, in that it attempts to reflect

the full set of social gains and social costs associated with existing human capital and the gains and costs of an investment in human capital.

Consider an individual at some point in her life, say age 16, who possesses some level of education, knowledge, and skills—human capital. By engaging in activities that contribute to the production of goods and services, she uses her human capital to produce outputs that are of value to the citizens of the nation, including herself. In the course of living and contributing to production, she uses up (or consumes) a variety of goods and services, and, therefore, the resources that are allocated to these outputs. Hence, she both employs her education and training in activities that contribute to social output, and in the process uses up resources that could be used to produce other things of value to society if they were not diverted to her. From the perspective of economic analysis, the value of her contributions to goods and services is measured by the full social opportunity costs associated with them.

Assume that for the current and each future year of her life we know both the value of what she contributes to society's output, and the value of social resources that she uses up or consumes. If we know the rates of time preference—or interest rates—of people who are positively and negatively affected by her activities, we can account for the fact that the value today of these future streams is less than if they were realized immediately. With this information, we can calculate the *present value* of the full lifetime stream of both the positive and negative contributions of her activities to social output.

The difference between the present value of her contributions to social output—call it her Gross Product—and the value of the social resources that she consumes is her net contribution to the nation, her Net Product.

Consider, now, that we are contemplating this same person but with an additional year of education, holding everything else about her constant. Given this framework, we can now pose the question of whether this additional year of schooling is a worthwhile social investment. The answer is clear: The additional investment is worthwhile if the person *with* the additional schooling has a greater Net Product than that same person *without* the additional year of schooling. In this case, the person's contributions to social output caused by the additional schooling exceed the social costs of providing those schooling services.

To see more clearly the nature of the gains associated with human capital and the costs required to produce human capital, it is helpful to decompose both the gain and cost components. Such a decomposition will more clearly reveal what is and what is not included in an assessment of the value of investments in education or human capital.

Table A1 is an annual statement of the production and consumption activities of the people who make up the society. The left side of the ledger tabulates contributions of these citizens to social output, and the right side calculates the value of society's resources that are consumed by the nation's citizens in any given year. Let us consider each of the two sides of the ledger in turn.

#### The Value of Gross Annual Product

The left side of the ledger itemizes the value of people's annual contribution to social output, the Value of Gross Annual Product. In making this tabulation, we adopt a comprehensive accounting stance, and include all of society's members.

Some of the activities of people yield contributions to the output of goods and services that pass through a market. Neglecting the complexities of self-employment, workers are likely to be employed by a firm and compensated for their labor effort. If the economy is a smoothly functioning market economy, the hourly wage is an estimate of the value of one hour's contribution to output; annual earnings (including fringe benefits) equal the value of the contribution to output for the entire year. This annual return reflects the knowledge and skills (human capital) that people possess and apply to market work during the year. We label this component the value of Market Production, or MP.

The logical underpinning of economics distinguishes an additional component of value beyond the market price of the goods and services produced. To the extent that people are willing to pay more than this market price, those purchasers of the goods and services realize a surplus. This Consumer Surplus (CS) is in addition to the value that the market places on goods and services produced. If the value of Market Production is measured using

Table A1

Value of Net Annual Product Balance She	et
Value of Gross Annual Product (VGAP)	Value of Annual Resource Use (VARU)
Value of Market Production (MP) [Often approximated by Earned Income (EI) (hourly market wage times hours engaged in Market Production) plus Fringe Benefits.]	Opportunity Cost of Food, Shelter, and Clothing Consumption (FSC) [Often approximated using market prices.]
Value of Home Production (HP) [Nonmarketed; often approximated by hourly market compensation times hours spent in Home Production.]	Opportunity Cost of Transportation and Medical Care Consumption (TMC) [Often approximated using market prices.]
Value of Volunteer Activities (VA) [Nonmarketed; approximated by hourly market compensation times hours spent in Volunteer Activities.]	Opportunity Costs of Education and Training Consumption (ET)
Consumer Surplus (CS) associated with Market Production (MP), Home Production (HP), and Volunteer Activities (VA)	MINUS Producer Surplus (PS) associated with FSC, TMC, and ET inputs when valued by market prices
Value of Leisure Activities (LA)	
Value of External Benefits (EB) Net Value to society, in excess of (MP + HP + VA + CS)	Value of External Costs (EC) Net value to society of costs in excess of (FSC + TMC + ET)

the market price of the output (as opposed to the full willingness of people to pay), the value of Consumer Surplus must also be included in the account. We enter it on the left side of the table after discussing the surplus values associated with other activities in which citizens engage.

The second entry on the left side of the ledger is the value of Home Production, labeled HP. In addition to productive activities that earn market rewards, citizens spend time in home-based work activities—caring for children, household maintenance, cooking, and numerous other tasks. These contributions to social output do not pass through a market, and people do not receive a monetary payment for doing them. Nevertheless, these contributions are as real as contributions that pass through a market; they also have value. Thus, a question arises concerning how to value such output.

Analysts often use an estimate of the market wage (including fringe benefits) that the person is (or would be) paid for Market Production as an approximation to the value to the individual of an hour spent in Home Production. The logic behind this reasoning is straightforward. If we assume, for a moment, that the individual allocates time over MP and HP, we know that each hour of MP earns her compensation equal to her market wage. Each hour of HP also grants "value" to her. If we further assume that each successive hour of HP grants less value than the previous hour, the individual will allocate hours to HP provided the value granted is greater than the market wage. Once the value granted from HP falls below the market wage, the individual will stop allocating hours to that type of production. Thus, one estimate of the value of Home Production is the hours spent in Home Production multiplied by the market wage rate (estimated if necessary and defined to include fringe benefits) to yield the aggregate value of home-based productive activities.

Of course, one implication of this reasoning is that producers of Home Production

receive value above their market wage rates. We will address this Producer Surplus in the next section. Here, however, we must address the Consumer Surplus associated with Home Production. Just as purchasers of marketed goods and services may be willing to pay more than the market price of those goods and services, "purchasers" of Home Production may be willing to pay more than the market price (now used as an estimate of the Home Production price) of Home Production. Here, HP fails to capture the implicit Consumer Surplus associated with home-based productive activities when it is measured using the market wage. We, therefore, separately account for this Consumer Surplus in Table A1.

After allocating time to the market and to the home, citizens have some time left for volunteer activities—time contributions to church, the local food pantry, neighborhood associations, school, and so on. The time that people spend in volunteer activities also yields services that are valuable to society, and again the appropriate concept for measuring the value of these services is the willingness to pay of all those who directly benefit from these outputs.

In practice, it is devilishly hard to approximate this value. Again, the hours do not pass through a market, and the value placed on them by the individual may be quite different than the value placed on them by society. However, as with home-based activities, analysts often equate the value of an hour of Volunteer Activities with the value of an hour of Market Production, again multiplying an estimate of hourly compensation by the number of hours citizens are engaged in Volunteer Activities. The logic is analogous to that used in the discussion of Home Production, extended to these activities. We enter the value of Volunteer Activities as the third item on the left side and label it VA.

Again, this estimate, based as it is on market values, neglects the implicit Consumer Surplus associated with Volunteer Activities. As we have noted, if the valuation of these productive human capital activities is based on prices reflected in the market, the estimated product will understate the full willingness to pay. To acknowledge this, we collect the Consumer Surplus values associated with MP, HP, and VA when valued by market prices, and include them in the left column of the ledger. They are labeled CS.

Beyond the hours not required for sleep and maintenance or used in these productive activities, people have residual hours of leisure that yield utility or well-being for themselves. Because each individual citizen is included as a member of society, the value of these leisure activities must also be tallied. The willingness-to-pay principle that guided the valuation of market, home-based, and volunteer activities also serves as the conceptual basis for valuing leisure hours. As with the other nonmarketed activities, analysts have attempted to use the expected market wage of people to approximate the value of their leisure hours. However, in this instance it is more difficult to make the case that people equate the market wage rate with the value of leisure hours, which is necessary for establishing the market wage rate as a reliable guide for valuing leisure hours. We include the willingness to pay for hours used in Leisure Activities as an entry in the table, and label it LA.

The last entry in the left column captures an important, but so far neglected, aspect of the value of the productive activities of citizens. To this point, we have assumed that the value of market, home-based, volunteer, and leisure activities can be secured from assessments of members of society (including the person whose human capital services are being valued) who directly benefit from these activities. In fact, these activities, particularly Home Production and Volunteer Activities, may increase the well-being of members of society who do not directly gain from the goods and services generated. For example, citizens in general may experience feelings of altruism (or "warm glow") when observing the benefits from the services of other citizens engaged in socially productive volunteer activities. This extra "spillover" or external value constitutes additional output, for example, in the form of better urban living conditions as a result of decreased homelessness, crime, or drug addiction, and must be included on our ledger. We label these external, public good-type benefits EB.<sup>21</sup>

The sum of the items in the left column of the ledger, then, is the annual social value

<sup>&</sup>lt;sup>21</sup> Although our examples indicate positive external effects, it should be noted that the productive activities reflecting the use of human capital may also generate negative effects. Hence, EB is appropriately thought of as a net value.

of the productive activities of citizens, with given education, training, skills, and other human capital characteristics. Because it captures the value of the services yielded by the human capital of citizens, without taking account of the social costs entailed in producing these outputs, this sum forms the gross annual return on human capital.

#### The Value of Annual Resource Use

Consider a unit of physical capital, such as a truck. For the truck to function productively, inputs for its operation, maintenance, and repair are required. In calculating the net value of the productive services of the truck to society, the analyst needs to take account of the value of these required inputs. The same is true of services rendered by people who embody human capital. Hence, we need a right side of the ledger to reflect the value of the annual resources diverted from other social uses in order to support and sustain the productive activities of human capital. These resources enable the person to live, work, and contribute the gross output indicated on the left side of the ledger.

Many of these inputs pass through a market; thus, valuing the opportunity cost to society in providing them to the individual is straightforward. However, the generation of these inputs for supporting human capital may also generate surpluse—in this case, producer surplus—that need to be taken into account in assessing these social opportunity costs. Moreover, the production and use of these goods and services may also impose external costs on society that are not reflected in market prices, and these costs must be included as well.

The primary required resources can be categorized in a rather straightforward manner; in each case it is the value of these inputs to society that must be assessed:

- Food, Shelter, and Clothing (FSC)—the basic necessities of life;
- Transportation and Medical Care (TMC)—other necessities with cost structures that are different from FCS;
- Education and Training (ET)—inputs supporting investments in human capital that will be used in productive activities in future periods;
- Producer Surplus (PS)—an offset to the market price of these required inputs, reflecting opportunity costs of productive factors that lie below market prices; and
- External Costs (EC)—nonmarketed costs generated in the process of producing these inputs to human capital, for example increased congestion or pollution.

The first entry on the right side of the ledger is the value of annual Food, Shelter, and Clothing (FSC) consumption by people. In concept, the social opportunity cost of these goods and services is the amount that would have to be paid to each unit of labor, land, and capital in order to divert it from some other activity into the production of FSC. A proxy measure of the opportunity cost of a unit of any one of these is its market price. Then the value of the annual resource use of these goods and services is the amount of each purchase multiplied by its market price.

If this market-based value is used to establish the value of FSC, social opportunity costs will be overvalued. Following the discussion of Consumer Surplus, we can argue that each successive unit of goods and services costs more to produce than the previous one, possibly because of higher labor costs or less efficient plants and equipment. However, the market price reflects the required cost to produce the last (or marginal) unit of these goods and services. If we value all the units produced at that market price, we overstate the total value of resources used. The magnitude of this overstatement is known as Producer Surplus, and must be subtracted from the total value of resources used on the right side of the ledger.

The second entry, Transportation and Medical Care (TMC), also reflects the value of inputs required for the productive use of human capital. As with FSC, the value of TMC is the social opportunity cost of the labor, land, and capital resources used in the production of these services, and analysts have made use of their market prices in developing proxies for the more difficult to measure, but conceptually accurate, social opportunity cost valuation. As we described above, such market prices tend to overstate the full social opportunity cost, by the amount of Producer Surplus; again an offset is required. However, in the case of TMC, market prices are far less reliable proxies of social opportunity costs than they are for FSC. Both medical care and transportation services enjoy public subsidies,

which lead to market prices that do not accurately reflect social costs. Hence, we include them separately in the ledger. The third entry, the value of Education and Training (ET) services consumed, represents the full social opportunity costs of the resources allocated to activities that augment the level of individual human capital stocks during a year. Unlike other real resource inputs required for productive activities that employ human capital-for example, consumption represented by FSC-the resources consumed for investments in human capital do not yield immediate increases in the value of productive activities that are reflected in the left side of this year's ledger. The added human capital stock will be put to productive use only in future periods, yielding gains in Gross Annual Product in these "out" years. For example, if the value of an hour of a person's contributions to Market Production (MP) is proxied by the hourly wage, the returns from the augmented human capital at the end of period t will be reflected in a higher hourly wage in future periods, implying increased market productivity in these periods. Because the value of a person's human capital stock is the discounted present value of the lifetime stream of her gross outputs, the gains that offset the value of the ET resource costs are reflected in the value of the human capital stock.22

As with TMC values, the market price of ET is a weak proxy for the relevant costs, due to public subsidies to both students and schools. And, as with FSC and TMC, Producer Surplus values will not be reflected in ET costs if market prices are used to assess the value of this resource use; again, these must be reflected as an offset on the right side of the ledger.

The next entry on the right side of the table is Producer Surplus, labeled PS. The value of PS is entered in the ledger with a minus sign, as it serves to offset the overstated costs of FSC, TMC, and ET when measured by market prices. As we have noted, if the valuation of the resources consumed in supporting the productive use of human capital is based on implicit or explicit market prices, the estimated value will exaggerate true social opportunity costs. Hence, we collect the Producer Surplus values associated with market-based estimates of these resource costs, and enter them on the right side of the ledger, but as an offset to the total resource cost.<sup>23</sup>

External Costs (EC), the final entry on the right side of the ledger, have the same conceptual basis as the value of External Benefits (EB) listed on the left side of the ledger. To the extent that those who bear the direct opportunity costs of the resources consumed in supporting the productive use of human capital do not experience the external or public goods costs of this consumption, they must be reflected in a separate entry in the ledger. An example of such costs borne by society but not directly reflected in the consumption of resources included in FSC, TMC, and ET are the pollution or congestion costs that may be associated with these uses of labor, land, and capital resources.

The sum of the items on the right side of the ledger is the annual social opportunity cost of the consumption of resources that support the productive activities associated with the use of human capital. When this sum is aggregated over all citizens in society it is the Value of Annual Resource Use (VARU) associated with the use of the human capital of the society.

We can now combine the two sides of the ledger. The gross annual value to society of the productive activities of human capital (the left side) minus the cost of the real consumption attributable to these activities (the right side) is the Value of Net Annual Product (VNAP) of human capital. VNAP is the value of the net annual contribution of

<sup>&</sup>lt;sup>22</sup> The decision to allocate time to Education and Training is more complicated than the decision to allocate time to other activities. The cost to society of an individual's choice to engage in education or training includes both the resource costs of the labor and capital inputs associated with the training, and the value of the individual's time devoted to that training in terms of lost output. Like the decision to allocate time to HP or VA, the individual will allocate time to ET as long as the value of that time to the individual exceeds the returns to time devoted to alternative uses. However, the time devoted to ET results in an increment to human capital, which in turn raises the individual's future productivity and wage rate and thus the value of all forms of the individual's future productive activities.

<sup>&</sup>lt;sup>23</sup> As with the value of Consumer Surplus, we include the Producer Surplus associated with the individual's own time spent in resource-using FSC, TMC, and ET activities.
human capital to aggregate output. It can also be considered the net annual social benefit of human capital, or the return on the stock of human capital existing at a point in time.

By extrapolating from this framework, then, we can define the annual social value of investing in one more unit of human capital—say, one more year of education for one person. The annual value of that investment is the *increase* in the Value of Net Annual Product of the society attributable to that choice, which equals the difference between the *increase* in the Value of Gross Annual Product and the *increase* in the Value of Annual Resource Use. The discounted present value of the full set of annual *increases* in the Value of the society attributable to that choice is the net social value of the investment. The social rate of return implied by the investment is the discount rate that would equate the present value of the full stream of increments to the Value of Gross Annual Product and the present value of the full stream of increments to the Value of Annual Resource Use.

## Discussion

## Social and Nonmarket Benefits from Education in an Advanced Economy

## Daron Acemoglu\*

The extensive literature on individual returns to education has been very influential on the thinking of economists and policymakers alike regarding the optimal amount of education for an individual and also the kinds of intervention that governments can or should undertake in educational markets. Wolfe and Haveman argue that many important social and nonmarket returns from schooling are being ignored. I am quite sympathetic to this view and believe that the investigation of nonmarket returns and externalities from education is a very important area for research.

Wolfe and Haveman focus on a number of these social and nonmarket benefits. In particular, they emphasize that greater schooling can lead to greater schooling of offspring, to better health for oneself and one's family members, to better consumer choices, to better fertility choices, to lower participation in criminal activities—and that greater schooling may have peer group effects related to the above choices.

This is a long list, and if only some of these benefits were important, it might be enough to change our views about what the optimal amount of schooling is for an individual from a social point of view. But one could add more standard social effects from schooling. There could be external returns to education. For example, the higher education of one's colleagues might increase one's productivity, or more-educated workers could undertake innovations that other workers in the economy might use. In addition, more-educated workers could make better political decisions.

Interesting related questions are whether these external and social

<sup>\*</sup>Professor of Economics, Massachusetts Institute of Technology.

nonmarket returns justify greater government intervention than we observe today, and whether these social returns have increased during the past 20 years as have private returns to education (private pecuniary returns to education), a phenomenon documented in inequality literature. Unfortunately, we do not get answers from Wolfe and Haveman.

My major concern with their paper is that despite the very important potential for new empirical work on these topics, the authors basically take a summary approach, and cite a large number of studies claiming these types of social and nonmarket effects. The problem is that all of these studies are ordinary least squares (OLS) estimates, which are driven by a variety of factors and do not establish that education, in fact, *causes* improvements in these various outcomes. Furthermore, the authors do not provide a satisfactory discussion of what these various effects actually mean, so it is difficult for the reader to understand what is an externality (that the government should care about) versus what is an effect that is already internalized by economic actors.

I find this of concern for two reasons: First, many of these effects may be present but may not correspond to any type of externality. For example, imagine that education leads to better consumer choices, but individuals are rational. Then when they are making their education choices, they take into account that not only will they earn more in the future, but also they will be able to get greater purchasing power from these wages because of their better consumer choices. In this case, the magnitude of these non-market effects is still useful to know for a variety of discussions, but there is no reason for the government to intervene, since these effects are already internalized. In other words, this type of discussion should start by a clear theoretical framework where we know what types of effects can be internalized, or are internalized in practice.

Second, and perhaps more serious, Wolfe and Haveman's paper takes existing associations in the data as the causal effect of education. It is quite possible that individuals who are more educated make better fertility choices or better consumer choices, but this does not mean that this is the causal effect of education on these choices. Individuals who obtain education are different, not only because of their ability, but also because of their parental and social background. It is quite likely that these background factors—not the education itself—lead to different consumer, fertility, or other social choices. These concerns lead to the question of what we actually know about any of these effects in a more careful empirical and theoretical setting.

Not surprisingly, here I would like to discuss some work that I have done on this topic, which explicitly deals with many of these issues. In joint work with Josh Angrist (2001), I investigated external effects to education in local labor markets. An often-expressed view, formalized among others in Acemoglu (1996), is that the productivity of workers increases when they are in the same labor market or in the vicinity of other more-educated workers. If true, this would be an important external effect from education, not internalized by individuals, and would provide a clear reason for government intervention to increase the education of workers throughout the economy.

To sort out these issues, the simplest strategy is to run a regression similar to the log-wage/education regressions that are very popular in labor economics, and add average education in the neighborhood or the local labor market of a worker. The following is an example of a simple regression of that form:

$$\ln w = X \cdot b' + a \cdot s + c \cdot S + e,$$

where w is the individual's wage, X is a vector of non-schooling attributes, s is own schooling, and S is average schooling in the same geographic location. For the purposes of this regression, the local labor market might be a city, a metropolitan area, or the state. Rauch (1993) has run this regression at the city level, and finds a very large coefficient on average schooling. Rauch interprets this as an external effect, arguing that workers receive higher wages, and most likely are more productive, when they are in the vicinity of other more-educated workers.

Josh Angrist and I ran the same regression at the state level, and similarly obtained a very large coefficient on average schooling. Moreeducated individuals receive higher wages, but also they tend to increase the wages of workers in the same labor market. In fact this OLS regression implies that the (local) externality is of the same magnitude as the private returns to education. An individual's own wages go up by 7 percent when he or she obtains one more year of education, but when the average education in the state increases by one more year, each individual's wages increase by an additional 7 percent—thus wages increase by a total of over 14 percent. Therefore, the external effect is an additional 7 percent on top of the private return of 7 percent. Does this then justify the conclusion that these are significant returns, and that there is room for more government intervention in educational markets?

No. As with all of the OLS studies, whether they are at the individual level or at the labor market level, there is a serious endogeneity problem. Cities with highly educated populations are different from each other in many aspects, including the amount of overall labor demand, and workers select into different cities or states based on their comparative advantage and abilities.

Putting state effects in a panel regression does not really solve these problems. We need a source of exogenous variation in the level of average education across various labor markets. Josh Angrist and I looked back to the early 1900s for big changes in compulsory schooling and child labor laws that affected various cohorts of individuals. Using these laws, we constructed instruments for individual and average schooling. We found that individuals growing up in states with tough child labor and compulsory schooling laws obtained significantly more education, and, as expected, this happens exactly at the point of dropping out of high school, not at the point of going to college.

Using this type of variation, which translates into substantial variation in average education across states at different points in time, we estimated the external returns to education. These instrumental-variables estimates paint a very different picture from the OLS estimates: There appears to be no evidence for large external effects. Our baseline estimates are around 1 percent and statistically not significant.

This evidence suggests that we should not rush to conclusions about the importance of external effects based on OLS evidence. This is somewhat more interesting for a personal reason, in that when I started the project I was convinced of the importance of external returns, based on my reading of the literature, case studies, and theoretical work that I had done previously (see Acemoglu 1996). However, once Josh and I became convinced (and managed to convince others) that we were exploiting the right type of variation in average schooling across states, the evidence was quite clear: no big externalities—in effect, no big \$500 bills lying on the street waiting to be picked up, even by the government.

Nevertheless, the absence of external returns in the labor market does not preclude the importance of other social and nonmarket benefits. A recent paper by Lochner and Moretti (2001) uses the compulsory schooling laws and the child labor laws that Josh and I put together to look at the effect of education on criminal activity. They find that individuals who obtain more education because laws prevent them from dropping out of school are less likely to commit a crime. This suggests that there might indeed be important nonmarket and social effects from education as argued by Haveman and Wolfe, though much more research needs to be done on the relationship between education and crime.

More generally, I think we should be looking for evidence of nonmarket and social effects from education in studies that are careful about the sort of variation used and that do not completely rely on association. We also need to start a serious discussion on the theoretical framework that distinguishes effects that are internalized for the individual versus effects that affect society as a whole, and thus can be properly be named "externalities." Overall, we have to thank Wolfe and Haveman for bringing this important issue back to the top of the agenda. There is a lot of exciting empirical and theoretical research awaiting us.

#### References

- Acemoglu, Daron. 1996. "A Microfoundation for Social Increasing Returns in Human Capital Accumulation." *Quarterly Journal of Economics* 111 (3): 779-804.
- Acemoglu, Daron and Joshua D. Angrist. 2001. "How Large Are Human Capital Externalities? Evidence from Compulsory Schooling Laws." NBER Macroeconomics Annual 2000, edited by B. S. Bernanke and K. Rogoff. Cambridge: MIT Press.
   Lochner, Lance and Enrico Moretti. 2001. "The Effect of Education on Crime: Evidence from Pricen Immetra. Arrosta and Coll Press." NDED W. Line Press.
- Lochner, Lance and Enrico Moretti. 2001. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." NBER Working Paper No. 8605 (November).
- Rauch, James E. 1993. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *Journal of Urban Economics* 3 (3): 380-400.

## Discussion

## Social and Nonmarket Benefits from Education in an Advanced Economy

## T. Paul Schultz\*

The objective of the paper by Barbara Wolfe and Robert Haveman is to assign a monetary value to the welfare gains associated with schooling that are not captured in the traditional measure of market-wage returns. Extending our perspective beyond the wage differentials of workers with different amounts of education promises to provide a more comprehensive basis on which to evaluate private and social investment priorities within the education sector, as well as the relative returns between education and alternative social investment sectors. This requires that we improve resource-accounting procedures and develop better methods to describe—without bias—the technologies that permit more-educated people to be more productive and thereby enjoy a higher standard of living.

Wolfe and Haveman note that if you can estimate without bias the parameters of the production function for nonmarket goods, which includes at least one market-priced input and one household worker's time input (distinguished by her or his education), then the trade-off between the marginal products of the worker's education and the marginal product of the market-priced input can be inferred. Thus they estimate a monetary value for the educational input, assuming that the ratio of the value of marginal products of all inputs divided by the inputs' market (or nonmarket) prices is equal when the allocation of inputs has been efficiently optimized.

To make this procedure more concrete, consider an example where an extra year of education of a mother increases her child's cognitive achievement by exactly the same amount as sending the child to a more

<sup>\*</sup>Malcolm K. Brachman Professor of Economics, Yale University.

expensive school. In this case, the added market cost of attending that better school is equivalent to a monetary valuation of the mother's production caused by her additional education in this one nonmarket production activity—child cognitive development. As Haveman and Wolfe note in their 1984 paper, several assumptions are required to justify their attractively simple methodology for estimating the social value of the many nonmarket private benefits and public externalities associated with education. Let me review some of the working assumptions that I think have proven unrealistic in subsequent studies, and thus represent limitations to their reported empirical findings.

The production of many goods and services that are consumed in the home by the producers and their families may be influenced by the education of family members in at least two distinct ways. Education may change the allocation of inputs and thereby increase outputs holding total costs constant. Alternatively, education can increase outputs, holding constant the mix of other inputs, presumably because a better-educated worker is intrinsically a more efficient producer with the same inputs in the same production process. Welch (1970) distinguishes between these two roles of education to clarify how a better-educated farmer managed to increase his profit and farm income, by both enhanced allocative efficiency and increased overall labor efficiency. In their analysis, Wolfe and Haveman allow for only the second pathway for education to impact social output by raising the overall efficiency of the worker's labor (per hour) and, consequently, they must assume "neutrality" of education in production. This implies that the Wolfe-Haveman approach is valid only if "the composition of other inputs does not change with changes in schooling" (Haveman and Wolfe 1984, p. 393). Is this a good approximation, given the limits of our measurement of nonmarket production technology? For farmers it is clearly a poor approximation (Huffman 2001).

I know of only one study of nonmarket production in which the effect of education on output is decomposed into education's production effect achieved via input reallocations, and via education, holding input allocations constant. In this study of a mother's production of birth weight as a proxy for child health, the production technology is hypothesized to be of a Cobb-Douglas form and uses four observed inputs in addition to education. The researchers could fully account for the significant positive partial association of mother's education and the expected birthweight of her child in the United States with the four input reallocations associated with the mother's educations associated with maternal education explained adequately the simple association of a mother's education of a mother's education and the improved birthweight outcome, leaving no significant residual to attribute to the "overall labor efficiency" effect of her education.

If this pattern is a typical or even a possible explanation for some of the partial associations between education and nonmarket productivity summarized in this paper, as I would anticipate if the technology is carefully dissected, one must ask, what is the cost of the reallocated input mix that the better-educated mother adopts? The Wolfe-Haveman methodology is likely to attribute all of the improved household production outcome to her education. I conjecture that her education may be associated with the use of a different, and probably a more expensive, mix of inputs in the home, many of which will not be observed in the typical survey or included in the regression analyses listed in the Wolfe-Haveman paper. The omission of such other household inputs as the mother's innate ability could also overstate nonmarket returns attributed to her education, for the same reason that many economists expect that the omission of ability in the analysis of wage determination might lead to an upward bias in the estimation of the private-market returns to education.

Education in such a production function could be viewed as more than a management capacity. Educated labor could also be an exhaustible resource, which must then be withdrawn from other valued activities at an opportunity cost to the family, in order to increase nonmarket production. That cost of the reallocation of family time is not discussed by Wolfe and Haveman, and of course it could be negative or positive. The more-educated mother may manage to produce healthier children while spending less of her time caring for her children. But other inputs are, in this case, likely to be substituted for her time, lowering the *net value* of her education's productive effect on child health, after deducting for the cost of those added inputs. I doubt that most of the studies of nonmarket private production cited by Wolfe and Haveman describe the determinants of the input used in household production and how education affects the use of all productive inputs, including the time of family members.

Heterogeneity bias is another general limitation for the study of household production functions (Rosenzweig and Schultz 1983), and is not discussed in the Wolfe-Haveman papers. The inputs to a household production function are allocated in response to *unobserved characteristics* of the individual producer. For example, the child's ability may affect which child goes to school; the initial health status of the child may determine whether the child receives medical care; the doctor's diagnostic ability may affect the health inputs the child is prescribed. The household's use of productive inputs is thus impacted by these variables that are unobserved by the researcher and that are likely to be associated with the residual variation in outputs—the child's productive capacity or the child's final health status, respectively. The estimates of the effects of the parents' education on household production will, in this case of heterogeneity of observations, tend to be biased and inconsistent, if estimated by single-equation methods, such as standard regression analysis employed in many of the studies cited by Wolfe and Haveman. One approach used increasingly to deal with this problem of heterogeneity is two-stage (least squares) or instrumental variable estimation methods. This involves predicting variation in input use related to variation in the exogenous instrumental variable, and this predicted variation in the inputs is then purged of the correlation with the residual variation in output from the household production function. Unbiased second-stage estimates of the household production function can thus be obtained, from which the marginal product of education can be inferred. A natural candidate for the instrumental variables in this context is the market-determined price of, or access to, the input that is observed to enter into the household production process.

Finally, the critical input that one wants to understand in the accounting exercise advanced by this paper is the reallocation of the time of the more-educated individual toward or away from the nonmarket production activity. Even if a mother today does not substantially reduce her time in market production to achieve an improvement in her child's health outcomes, she may sacrifice her leisure as well as use other market goods and services to substitute for her own home childcare time. The exercise reported in the Wolfe-Haveman paper is, therefore, only the starting point for a more comprehensive evaluation of the net benefits arising from education contributing to enhanced nonmarket production. In other words, from the gross output association reported in their paper, one must subtract the opportunity cost of lost leisure and market work time, if any, and the market cost of other inputs substituted for the time of the more-educated mother in the nonmarket production process. Therefore, I would conjecture that the nonmarket private productive values of education reported in this paper are substantially upward biased. Improved research is now needed to confirm this intuition.

As the authors note at the start of their paper, the economics literature on the market-wage returns to schooling as approximated in a Mincerian wage function has evolved through a variety of conceptual and statistical interpretations in the last 40 years. Literally hundreds of empirical and statistical papers were produced before there emerged a consensus that the widely anticipated "ability bias" that is expected to overstate the wage returns to schooling is not in fact all that substantial. The first generation of studies summarized by Griliches (1977) concluded that the errors in measuring education were an offsetting source of downward bias, which apparently canceled out the upward ability bias. In the second generation of work, summarized by Card (2001), the application of instrumental variable estimation methods, which exploit the variation in school-system supply factors for identification—in other words, building schools in the neighborhood of the respondent—has yielded somewhat higher estimated wage returns than implied by the simple regression approach pioneered by Jacob Mincer. Card concludes that this upward adjustment in schooling returns associated with supplybased instruments may be due to the heterogeneity of the labor productivity gains realized from schooling. He hypothesizes that the expansions in public supplies of school service disproportionately benefit disadvantaged children from credit-constrained families and that these children may earn higher-than-average returns on their education.

This estimation methodology must now be extended to grapple with the task of estimating nonmarket production functions without bias, in order to proceed to answer the question addressed in the Wolfe-Haveman paper. A similar problem arises from the heterogeneity of students and schools, which affects the endogenous allocation of inputs among and within schools and contributes to bias in the estimation of educational production functions by ordinary single-equation methods. Interpreting with caution the existing empirical evidence derived from educational production-function studies of schools is imperative, except when input or process variation is implemented randomly across treatment and control populations. (This important issue is relevant to the Hanushek and Raymond paper included in this volume.)

This history of statistical and empirical studies of the market-wage returns to education is well known, but I have restated it here to underscore the point that to improve our empirical knowledge of the private market-wage returns to education, many increasingly sophisticated studies were undertaken that explicitly allowed for the heterogeneity of people and their environments, and in many cases natural and social experimental settings were exploited with and without instrumental variable methods. There is now a need to improve the first-generation studies of nonmarket household production, which are reviewed in the Wolfe-Haveman paper, because most of these studies ignore these problems. A more satisfactory statistical and conceptual approach to household production (one that recognizes that the education of family members influences the mix of inputs used in the household, and changes the allocation of the time of household members) will be needed. Because of the heterogeneity of households and their endogenous choice of inputs, including their own time allocations, the reported empirical evidence of private nonmarket returns to education as reported by Wolfe and Haveman is not satisfactory as even a first approximation, although I have no doubt that nonmarket privately realized returns in the formation of children's human capital are probably substantial. But these are privately captured returns, if altruistic parents benefit from these returns, and thus not commonly viewed as a rationale for public sector subsidies for education.

I look forward to a new generation of empirical research into the role of education in household production, from which more adequate and less biased evaluations of the nonmarket returns to education can be derived using the conceptual logic outlined in the Wolfe-Haveman paper. I anticipate that these new studies will show substantial returns in welfare improvements beyond market-wage returns to schooling, which arise from the enhanced production of nonmarket outputs, such as the health and schooling of children. But we do not yet have these studies in hand, and the remaining task is not a trivial one, in terms of collecting suitable data and their correct analysis.

In conclusion, I should indicate that I have commented only on the private nonmarket returns to education. I have not elaborated on the equally daunting analytical challenges that face Wolfe and Haveman if they want to develop more satisfactory estimates of the social externalities from education that spill over beyond the family. In his comments on this paper, Daron Acemoglu reviews these issues involving the definition and measurement of social externalities of education. (See also Acemoglu and Angrist 1999; Moretti 1998.)

#### References

- Acemoglu, Daron and Joshua Angrist. 1999. "How Large Are the Social Returns to Education? Evidence from Compulsory Schooling Laws." NBER Working Paper No. 7444 (December).
- Card, David. 2001. "Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69 (5): 1127-61.
- Griliches, Zvi. 1977. "Estimating the Returns to Schooling: Some Econometric Problems." Econometrica 45 (1): 1-45.
- Haveman, Robert H. and Barbara L. Wolfe. 1984. "Schooling and Economic Well-Being: The Role of Nonmarket Effects." *Journal of Human Resources* 19 (3): 378-407.
- Huffman, Wallace E. 2001. "Human Capital: Education and Agriculture," in *Handbook of Agricultural Economics*, edited by B. L. Gardner and G. C. Rausser, vol. 1A. North Holland: Elsevier, Amsterdam.
- Moretti, Enrico. 1998. "Social Returns to Education and Human Capital Externalities: Evidence from Cities." Center for Labor Economics Working Paper No. 9. University of California, Berkeley (December).
- Rosenzweig, Mark R. and T. Paul Schultz. 1982. "Education and the Household Production of Child Health." 1981 Social Statistics Section Proceedings of the American Statistical Association. Washington, DC.
  - —. 1983. "Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and the Effects on Birthweight." *Journal of Political Economy* 91 (5): 723-46.

Welch, Finis. 1970. "Education in Production." Journal of Political Economy 78 (1): 35-59.

# **Do State Governments Matter?**

#### A REVIEW OF THE EVIDENCE ON THE IMPACT ON EDUCATIONAL OUTCOMES OF THE CHANGING ROLE OF THE STATES IN THE FINANCING OF PUBLIC EDUCATION

## Thomas A. Downes\*

During the past several decades, federal and state governments have pursued a variety of redistributive policies aimed at fostering the idea of "equality of economic opportunity." This concept implies that although people's incomes may vary, the variance should be caused by factors such as individual ability and effort, not by differences in circumstance. Many in the policy arena have suggested that opportunities could be further equalized by implementing changes in the way elementary and secondary education is financed and delivered. Hanushek and Somers (1999) detail the most prominent state and federal policy initiatives aimed at reducing income inequality by modifying education finance and delivery.

This paper focuses on three sets of changes to the school finance landscape, and attempts to summarize the evidence on the effects of these changes on education quality and, ultimately, on the extent of inequality in American society. The first set of changes considered will be school finance reform and the large-scale changes in the formulas states use to determine aid to local school districts. For many years, those concerned with the persistence of income inequality in the United States have argued for reforms to the method of financing public elementary and

<sup>\*</sup>Associate Professor of Economics, Tufts University and Visiting Scholar, Federal Reserve Bank of Boston. The author thanks Sheila Murray, Larry Kenny, Julian Betts, Eric Hanushek, and participants in the Federal Reserve Bank of Boston's conference, "Education in the 21st Century: Meeting the Challenges of a Changing World," for their helpful comments and suggestions.

secondary schools that would make education spending more equal. These arguments, which have been buttressed by substantial evidence that pre-market factors play a significant role in determining subsequent labor market outcomes (see, for example, Murnane, Willett, and Levy 1995; Neal and Johnson 1996; Bishop 1989), have been cited by those who have argued in the courts for fundamental reforms of the way in which public schools are financed (see, for example, Campaign for Fiscal Equity 2001). These court challenges have experienced a resurgence in the last several years, with state supreme court decisions mandating equalization in states such as Kentucky, Texas, Vermont, and New Hampshire further altering a school finance landscape that has changed dramatically since 1970. The end result of these court challenges is that, in almost every state in the nation, the system of financing the public schools has been fundamentally altered, with state governments becoming an ever more important part of the educational financing landscape. In this paper, I review the empirical evidence on the effects of these changes, concentrating on the relationship between school finance reforms and student outcomes.

The second set of changes summarized in this paper are those attributable to tax and expenditure limitations. Limitations on the ability of local governments to raise revenues or to make expenditures, like those imposed by Proposition 13 in California and Proposition 21/2 in Massachusetts, have typically forced state governments to increase state-level taxes and state aid to public schools (Galles and Sexton 1998; Cutler, Elmendorf, and Zeckhauser 1999). From a distributional perspective, the effects of tax limitations on school accountability efforts are also important. I make no attempt to evaluate high-stakes testing and other existing accountability measures; I will leave that for other papers in this volume. Nevertheless, I will review evidence that suggests that tax limitations have significant imbedded incentives that may result in outcomes that are different from the intended consequences of increased fiscal accountability. Many of these incentives are present in other types of accountability systems, so it would behoove today's policymakers to take to heart the lessons of the tax revolt.

The third set of changes I address in this paper involves school choice. While no state has implemented the type of voucher system that Friedman (1955) advocates, small-scale, publicly and privately funded voucher plans exist in several localities. Further, 37 states and the District of Columbia currently have charter school laws, all of which, to a greater or lesser extent, allow for increased choice within the public system. In this paper, I will limit myself to recounting the evidence on the impact of the charter school movement, since this movement has been driven by state-level policy changes and because it represents the most widespread challenge to the traditional system of school finance. Documenting the

effect of charter schools will, over the next few years, be one of the most important tasks for researchers.

Each of the three changes to the school finance landscape could have numerous effects beyond student performance, which is the focus of the research summarized in this paper. Take, for example, school finance reforms. Researchers have attempted to quantify the effect of these reforms on house values (Dee 2000), community composition (Aaronson 1999; Downes and Figlio 1999b), private school attendance (Downes and Schoeman 1998), private school supply (Downes and Greenstein 1996), and private contributions to public schools (Brunner and Sonstelie 1997). My decision to restrict my discussion to the impact on student performance is driven by two considerations. First, policymakers are particularly interested in the impact of policy changes on student performance on standardized tests. This is made readily evident by the increasing prevalence of high-stakes testing and by the provisions in the No Child Left Behind Act of 2001 that make federal aid contingent upon measurable improvement in the quality of services provided. Second, as Hanushek and Somers (1999) note, recent research has documented a strong link between standardized test scores and earnings. Thus, policies that reduce dispersion in standardized test scores should, ultimately, reduce dispersion in earnings. Nevertheless, I do not want to leave the reader with the impression that the impact of these policy changes on the distribution of some of the other determinants of social well-being that are catalogued by Wolfe and Haveman (2002) is either uninteresting or unimportant. Quantifying these impacts unquestionably will be neces-sary in order to estimate the overall welfare implications of these policy changes.

The absence of good measures for many of the determinants of social well-being will make it difficult to quantify the link between these policy changes and the distribution of social well-being. But even quantifying their effect on the distribution of student performance, which is easily observed, has proven to be challenging, since none of these policy changes has occurred in isolation. One of the problems researchers have had to overcome is how to isolate the impact of one change in the system of school finance and delivery from all of the other changes, large and small, that are being implemented contemporaneously. Improvements in available data and in econometric techniques have, in recent years, resulted in an increasing number of studies attempting to isolate the effects of these policy changes on the distribution of student performance. Accompanying this literature have been several papers that critically review the literature. I have drawn heavily from these reviews, and I strongly encourage the interested reader to turn to these reviews for more exhaustive summaries of the existing state of knowledge. For school finance reforms, Murray (2001), Downes and Figlio (1999b, 2000), and Card and Payne (2002) offer alternative views of the effects. Downes and Figlio (1999a, 2001) and Kirchgässner (2001) summarize the literature on the impact of tax limitations. And the recent monograph authored by Gill et al. (2001) represents a thorough, careful, and dispassionate overview of the evidence on the effects of voucher programs and charter schools; Miron and Nelson (2001) summarize and critically evaluate much of the research on the impact of charter schools on student achievement.

There are no pithy remarks that can summarize the lessons from this paper. The reality is that, while much progress has been made on quantifying the effects of these changes in the school finance landscape, much work remains to be done. Only for tax and expenditure limitations has any consensus concerning their effects on mean achievement begun to develop, and even for tax and expenditure limitations there is much still to be learned about their distributional implications. The main lesson, then, is that there is considerable room for additional research into the achievement effects of each one of these sets of policies.

The next section of the paper summarizes research examining the links between school finance reforms and student achievement. A review of the evidence on the effects of tax and expenditure limitations follows, as does a very brief discussion of the implications of this evidence for other accountability measures. The limited work on the impact of charter schools on student achievement is then presented. The paper closes with some suggestions for future research.

### A REVIEW OF RESEARCH ON THE IMPACT OF SCHOOL FINANCE REFORMS

The school finance reforms implemented in California after the *Serrano* v. *Priest* case and Proposition 13 represent a watershed both in the debate over the structure of school finance reforms and in the direction of research into the impact of those reforms. In the post-*Serrano* period, the California reforms and their supposed effects on the schools in that state have been discussed in every state in which school finance reforms have been implemented.<sup>1</sup>

The California reforms also shifted the focus of research to the impact of school finance. Prior to the reforms, the focus in the literature was almost solely on the impact of finance reforms on spending inequality. After *Serrano*, the scope of the analysis broadened to include the impact of finance reforms on the level and distribution of student achievement, on housing prices, on the supply of private schooling, and

<sup>&</sup>lt;sup>1</sup> For instance, in Vermont, where Act 60 represents the most radical of the recent school finance reforms, examples of references to California include McClaughry (1997) and Mathis (1998).

even on the composition of affected communities.<sup>2</sup> The California reforms also became the touchstone for theoretical work. Papers like those of Nechyba (1996, 2000), Bénabou (1996), and Fernandez and Rogerson (1997, 1998) use a California-like system as the post-reform case when trying to reach predictions about the likely effects of finance reform.

The problem with using the California case as a benchmark is that the case has proven to be the exception, not the rule. First, the limits imposed on local control over spending have not been duplicated in any other state. Even in Michigan and Vermont, the states in which the most extensive post-*Serrano* reforms have been implemented, some degree of local control over taxes and spending is permitted. Further, the population of students served by California schools changed more dramatically than the population of students in any other state in the nation. From 1986 to 1997, the proportion of the California public school student population identified as minority increased from 46.3 percent to 61.2 percent. Nationally, the minority share grew far more slowly, from 29.6 percent to 36.5 percent. As Downes (1992) notes, these demographic changes make it difficult to quantify the impact of the finance reforms in California on the cross-district inequality in student achievement.<sup>3</sup>

The possibility that California might be the exception and not the rule pushed a number of researchers to pursue national-level studies attempting to document the impact of finance reforms. On the spending side, Silva and Sonstelie (1995), Downes and Shah (1995), and Manwaring and Sheffrin (1997) each take slightly different approaches to quantifying the effect of finance reforms on mean per pupil spending in a state. Because they use district-level data, Hoxby (2001a), Evans, Murray, and Schwab (1997), and Murray, Evans, and Schwab (1998) are able to consider not only the effects of finance reforms on mean spending but also the extent to which spending inequities were reduced by those reforms. As a result, these studies provide the most obvious sources for predictions of the long-run effects of school finance reforms.

The problem is that these studies generate contradictory predictions. The case of Act 60 in Vermont helps make concrete the disparity in predictions. Hoxby's results would lead us to expect leveling down, since Act 60 dramatically increases tax prices in towns with more property

<sup>&</sup>lt;sup>2</sup> The papers dealing with these varied topics are too numerous to cite. Evans, Murray, and Schwab (1997) and Downes and Figlio (1999b, 2000) cite many of the relevant papers.

<sup>&</sup>lt;sup>3</sup> Generating comparable numbers for earlier years is difficult. Nevertheless, the best available data support the conclusion that these sharp differences in trends in the minority share pre-date the *Serrano*-inspired reforms. For example, calculations based on published information for California indicate the percent minority in 1977–78 was approximately 36.6 percent. Nationally, estimates based on the October 1977 Current Population Survey indicate the percent minority was 23.9 percent.

wealth. Murray, Evans, and Schwab conclude that court-mandated reforms like Act 60 typically result in leveling up.

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

While the dropout rate is an outcome measure of considerable interest, analyses of the quality of public education in the United States tend to focus on standardized test scores and other measures of student performance that provide some indication of how the general student population is faring. Husted and Kenny (2000) suggest that equalization may detrimentally affect student achievement. Using data on 37 states from 1987–88 to 1992–93, they find that the mean SAT score is higher for those states with greater intrastate spending variation. However, the period they consider post-dates the imposition of the first wave of finance reforms. Thus, the data do not permit direct examination of the effects of policy changes. In addition, because they use state-level data, Husted and Kenny cannot examine the degree to which equalization affects cross-district variation in test scores. Finally, since only a select group of students take the SAT, Husted and Kenny are not able to consider how equalization affects the performance of all students in a state.<sup>4</sup>

Card and Payne (2002) explore the effects of school finance equalizations on the within-state distributions of SAT scores. They characterize a school finance policy as more equalizing the more negative is the within-state relationship between state aid to a school district and school district income. They find that the SAT scores of students with poorly educated parents (their proxy for low income) increase in states that, under their definition, become more equalized. Data limitations, however, make it impossible for Card and Payne to examine the effects of

<sup>&</sup>lt;sup>4</sup> Husted and Kenny do find evidence consistent with the conclusion that, in states in which school finance reforms had reduced the dispersion in per pupil expenditures, these reforms have had no impact on the standard deviation of SAT scores. Since, however, the standard deviation of test scores could be unchanged even if cross-district inequality in performance had declined, this evidence fails to establish that finance reforms do not reduce cross-district performance inequality.

policy changes on students residing in school districts in which the changes had the greatest impact. Moreover, while Card and Payne correct for differences in the fractions of the population taking the SAT, it is still very likely that the students who come from low-education backgrounds but take the SAT are a very select group and are extremely unlikely to be representative of the low-income or low-education population as a whole.<sup>5</sup>

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

Downes and Figlio (2000) find that finance reforms, in response to court decisions, result in small and frequently insignificant increases in the mean level of student performance on standardized tests of reading and mathematics. Further, they note that there is some indication that the post-reform distribution of scores in mathematics may be less equal. This latter result highlights one of the central points of the paper: Any evaluation of finance reforms must control for the initial circumstances of affected districts. The simple reality is that finance reforms are likely to have differential effects in initially high-spending and initially lowspending districts.

Downes and Figlio's (2000) finding that court-ordered finance reforms may be associated with increased dispersion in student performance is echoed by results produced by Hanushek and Somers (1999). Hanushek and Somers use data on earnings of workers who are 25 to 37 years old in 1990 to calculate within-cohort variation in earnings. Like Husted and Kenny (2000), they do not directly estimate the effect of finance reforms, choosing instead to relate the extent of equalization in a state to the extent of earnings variation among those who were born in that state. They find that, for white males and females, earnings variation

<sup>&</sup>lt;sup>5</sup> For instance, among the students in Card and Payne's low-parental-education group, in 28 states in 1978 (25 states in 1990) fewer than 10 percent took the SAT examination and in 20 states in 1978 (15 states in 1990) fewer than 3 percent took the SAT. Further, in 1978 no state had more than 36.2 percent of the low-parental-education group take the SAT.

is negatively related to the extent of spending variation across high schools in the cohort's birth state at the time the cohort attended high school. Only for black females is there any evidence that reductions in school spending variation are associated with reductions in earnings variation.

The contrast between the results of Card and Payne and those of Hanushek and Somers highlights the challenge facing anyone trying to predict the impact of potential reforms on any state's system of school finance. The fundamental reason for the absence of clear predictions of the impact of finance reforms has been mentioned by a number of authors (see for example, Downes and Shah 1995; Hoxby 2001a; Evans, Murray, and Schwab 1997), all of whom have emphasized the tremendous diversity of school finance reforms. In a national-level study, any attempt to classify finance reforms will be imperfect. Even though there is general consensus that the key elements of a finance reform are the combined effects of the reform on local discretion and on local incentives, and the change in state-level responsibilities in the aftermath of reform (Hoxby 2001a; Courant and Loeb 1997), different authors take different approaches to account for the heterogeneity of the reforms. The result is variation in predictions generated by studies that are asking the same fundamental question. The answer is not, it seems, to try to improve the methods of classifying reforms but is, instead, to complement these national-level studies with case studies of canonical reforms. Only national-level studies can reveal if, in a state in which school finance reforms have been implemented, the mean performance of students has changed relative to what this performance would have been in the absence of the finance reforms. If, however, the research question is whether the finance reforms have altered the distribution of student performance, both state-level and national-level studies can provide results that can be used to answer the question. And only state-level case studies can convincingly indicate which, if any, characteristics of reforms are linked to success in reducing the extent of performance inequality.

The most direct antecedent in this case-study approach to analyzing finance reforms is Downes (1992), who shows that the extensive school finance reforms in California in the late 1970s generated greater equality across school districts in per pupil spending but not greater equality in measured student performance. For all the reasons noted above, replicating this style of analysis for other states is imperative. Downes's (2002) work on Vermont, Flanagan and Murray's (2002) work on Kentucky, and Duncombe and Johnston's (2002) work on Kansas offer examples of recent case studies of canonical reforms.

The diversity of school finance reforms is apparent as one looks across these case studies. What is striking is the similarity across studies in the estimated achievement effects. Pre-finance reform data on student test scores are not available to Duncombe and Johnston; they find no evidence of diminished dispersion in performance when examining post-finance-reform test scores. They also document some recent relative improvement in dropout rates in high-poverty districts, though they also find increased dispersion in dropout rates when comparing pre- and post-finance-reform data.

The bottom line of Duncombe and Johnston's analysis of dropout rates is that reform has resulted in small relative improvements. Downes (2002) and Flanagan and Murray reach similar conclusions-post-reform dispersion in schooling outcomes has declined, but this decline in dispersion has been small. Downes finds that there have been, at most, small relative improvements in the test performance of fourth and eighth graders in those school districts with lower pre-reform per pupil spending and per pupil property wealth. Flanagan and Murray find that relative increases in post-reform spending were translated into relative gains in post-reform test performance, but these gains were quantitatively small. Somewhat surprisingly, then, the results of these new case studies tend to echo the results of the earlier work on California. Thus far, the case studies have confirmed a conclusion that was reached by many of the researchers who executed national-level analyses: The types of finance reforms that have been implemented in response to court orders appear to have little, if any, impact on the distribution of student test performance.

# Do Tax and Expenditure Limits Affect Student Performance?

Like research into the impact of school finance reforms, research into the effects of tax limits blossomed after a major policy change in California. Much of this research focused, however, on the fiscal implications of tax limits; see Fisher (1996) for an interpretive review of this work. The passage of Proposition 13 in California in 1978, followed by the 1981 approval of Proposition 21/2 in Massachusetts, did not stimulate immediate research on the impact of tax and expenditure limits on student performance. That there was a lag between implementation of these limits and research on the link between the limits and schooling outcomes is not surprising, since limits are unlikely to affect the performance of most public students in the short term. What is surprising is that by 1990 there were few studies in which the impact of limits on service provision was examined. Further, the studies that existed were exclusively case studies that considered the effects of limits on a variety of services, including public education, but that did not use explicit measures of student performance to gauge the effects of limits on those served by the schools.

Nevertheless, case studies like those of the Joint Budget Committee (1979) and Schwadron (1984) for California and Greiner and Peterson

(1986) for Massachusetts present a relatively consistent picture of the short-run effects of tax limits on service quality. In general, residents of the states considered by the referenced studies perceived a drop in service quality. That this perception reflected reality was sometimes, though not always, confirmed by objective measures of service quality (Greiner and Peterson 1986). Government officials responded to the limits by first making cuts in capital expenditures and in areas of current expenditure that these officials felt were peripheral. For example, in California, school administrators sought to protect the core academic subjects, choosing instead to make cuts by pursuing such strategies as reducing the diversity of course offerings and the number of pupil service employees.<sup>6</sup>

Given their timing, these case studies could not be used to draw any conclusions about the long-run effects of tax and expenditure limits. Also, even though these case studies moved beyond examination of the fiscal impacts of limits, the concerns raised in the introduction to this paper imply that the results of these case studies could not be used to predict with confidence the effect of limits on student outcomes. Only by examining student outcomes directly and by determining how these outcomes had changed relative to the pre-limit baseline, could researchers ascertain the effect of limits.

The first research to compare pre- and post-limitation measures of student performance was Downes (1992). In that study, data on district means of performance on the California Assessment Program test were assembled for 170 unified (K–12) districts in 1976–77 and 1985–86. In these districts, the measure of student performance actually increased by 5 points, on average. Further, the cross-district distribution of student performance was essentially unchanged between 1976–77 and 1985–86. The bottom line of this research, it would seem, is that Proposition 13 did not produce a long-run reduction in student performance at any point on the performance distribution.

Such a conclusion would be unwarranted, however. As was noted above, contemporaneous with the state and local response to Proposition 13 was state implementation of school finance reforms made necessary by the *Serrano* decision.<sup>7</sup> This observation raises a problem that faces any researcher attempting to isolate the impact of tax limits on public

<sup>&</sup>lt;sup>6</sup> The results in Downes (1996) suggest that school administrators in California did not respond to the limits by cutting the administrative staff. For a national cross-section, Figlio (1997) also finds no evidence of cuts in administration.

<sup>&</sup>lt;sup>7</sup> Fischel (1989, 1996) makes a strong case that, in fact, the prospective school finance reforms that were compelled by the *Serrano* decision stimulated enough additional support for tax limits to make passage of Proposition 13 inevitable. If this logic is right, any observed changes in the distribution of student performance in California should ultimately be attributed to the finance reforms, not the resultant tax limits.

schooling. Frequently, states have implemented major school finance reforms close in time to the passage of tax limits. Thus, the effects of either school finance reforms or tax limits can be isolated only by looking across states or by examining the long-run experience in a state in which a limit was passed and no major changes in the school finance system had occurred.

Three recent papers take this lesson to heart and, thus, provide a model for future empirical research on the impact of tax and expenditure limits. Using a cross-section of student-level data from the National Education Longitudinal Survey (NELS), Figlio (1997) finds that, all things equal, the performance of tenth graders on mathematics, reading, science, and social studies tests was significantly lower in those states in which local school districts faced either revenue or expenditure limits. Since, however, Figlio's variation was cross-sectional, he was unable to rule out the possibility that some combination of sorting and unobserved tastes for education resulted in both the passage of limits and less rapid improvement in student performance.<sup>8</sup>

To avoid this problem, Downes, Dye, and McGuire (1998) examine the recent imposition of property tax limits on school districts in the Chicago suburbs. They conclude that, in the short term, these limits translated into slower growth in the performance of third graders on a standardized test of mathematics. Similar slowing of growth is not observed for third-grade reading test scores or for the test scores of eighth graders. The authors also note that the effects of these limits varied across districts. What the authors could not do is argue that their results provide a definitive picture of the long-term effects of any tax or expenditure limits, since they observe only three post-limit years and since the Illinois case could be exceptional. Their paper is difficult to draw conclusions from because, like Figlio's paper, Downes, Dye, and McGuire's conclusions are driven by unobserved differences between the "control" and "treatment" groups in the analysis.

The third paper, Downes and Figlio (2000), discussed earlier, builds on the strengths of these two studies. One lesson from these two papers is that evaluating the effects of tax limits requires not only before and after data on students in districts subject to limits but also a control group of students from states in which no limits have been enacted. With this observation in mind, Downes and Figlio attempt to determine how the tax limits of the late 1970s and early 1980s affected the distribution of

<sup>&</sup>lt;sup>8</sup> The same problem plagues the work of Shadbegian (2001), who finds that student test performance is lower in those Massachusetts districts forced to cut property taxes in the aftermath of Proposition  $2\frac{1}{2}$ . Unfortunately, because he has no data on pre-Proposition  $2\frac{1}{2}$  test performance, Shadbegian is unable to rule out the possibility that there exist unobservable factors that resulted in lower test performance and that are correlated with the extent to which a locality was constrained by Proposition  $2\frac{1}{2}$ .

student performance in states in which limits were imposed and how student performance has changed in these states relative to student performance in states in which no limits were imposed.

The results in Downes and Figlio confirm, in part, the results of Figlio (1997). Specifically, the imposition of tax or expenditure limits on local governments in a state reduced student performance on standardized tests of mathematics skills by 1 to 7 percent, depending on model specification. However, there was no general evidence that tax limits affected student performance on standardized tests of reading skills, except when tax limits were treated as endogenous—that is, when the researchers estimated a regression model in which the possible reverse causality between test scores and tax limits was taken into account.9 This latter result-no general finding of an effect on reading performanceparallels one of the findings of Downes, Dye, and McGuire. It is sensible, given the age of the test-takers, to believe that high school mathematics differences may be more attributable to differences in schooling than are high school reading differences, so the generally stronger effect of tax limits on mathematics than on reading should not come as much of a surprise.

For the most part, when researchers have examined the impact of tax limits on student performance, they have confined their analysis to students who remain in the public schools. Bradbury, Case, and Mayer (1998) represents a break from this norm, analyzing the relationship between grade-level enrollment patterns and various indicators of the bindingness of tax limits. Since differences between actual enrollment patterns and the patterns of enrollment implied by the decennial Censuses reflect primarily withdrawal from the public schools, either to private schools or nonenrollment status, the results from their paper shed some light on the effect of tax limits on dropout rates. Bradbury, Case, and Mayer find that the share of the potential student population served by the public schools is lower in districts in which more initial cuts were necessary when the limits were first imposed. This result suggests that limits could increase dropout rates, though further research on this question is clearly needed.

Another recent paper, Downes and Figlio (1999b), provides the first attempt to study the performance effects of tax limitations (and school finance reforms) on private school students. This study uses a similar methodology to that used by Downes and Figlio (2000) to investigate the effects of tax limits on public school performance. While their results are more compelling for school finance reforms than for tax limitations, Downes and Figlio (2000) find limited evidence of a modest (though

<sup>&</sup>lt;sup>9</sup> See Figlio (1997) or Downes and Figlio (2000) for more of a discussion of the potential endogeneity biases, as well as a detailed treatment of the issue of reverse causality.

imprecisely estimated) negative effect of tax limits on student test scores in the private sector. This result, if one considers only the magnitude and not the statistical significance of the finding, could be interpreted in several ways. One possibility is that tax limits may tend to lower the quality of the private sector, either because of lower competition from the public sector or for other reasons, such as peer effects. Another possibility is that the lower test performance is a manifestation of increased selection into the private sector by students less able than those who populated the private sector before the limitations' passage (though still, on average, more able than the typical public school student).<sup>10</sup> Though this line of research provides a first look at the overall distributional consequences of tax limitations, it is clear that much more work is needed on this topic.

Evidence on the impact of tax and expenditure limits on the cross-district distribution of student performance, while consistent across studies, is less compelling than evidence on the impact of these limits on mean performance. Specifically, Downes, Dye and McGuire (1998) and Downes and Figlio (2000) find that student performance appears to deteriorate more in economically disadvantaged localities, though these cross-locality differences—while consistent in direction across specifica-tions—frequently proved to be statistically insignificant. Nevertheless, this limited evidence on the nonuniformity of the effects of tax limits suggests the need for further research on the dependence of these effects on a district's initial conditions and demographics.

While it is not clear whether tax limitations are good policy, arguably this literature does clarify that policies with one set of desired outcomes may have another set of unintended consequences—both favorable and unfavorable. These lessons are interesting on their own merits, but they are also important because of the possible applicability of these lessons from the fiscal accountability-driven tax revolt to the new wave of public accountability.

Since the early 1990s, there has been a national trend toward increased school-level accountability in education. Today, almost every state in the United States conducts regular testing of students, and most have high stakes attached to student test performance, such as potential grade retention or failure to graduate from high school. Because these accountability policies are so new, there has been virtually no formal evaluation of their effects. However, we know from the literature on tax and expenditure limitations that one possible reason for reductions in

<sup>&</sup>lt;sup>10</sup> Epple and Romano (1998) theoretically describe stratification patterns between the public and private sectors that predict precisely this result—that reduced public sector spending leads to the movement of "top" public school students into the private sector, reducing the average performance level of both the private and public sectors. Epple, Figlio, and Romano (1998) offer some empirical justification of the stratification patterns identified in the theoretical model.

student performance in excess of what might be expected given the change in financial resources is that the incentives associated with tax limits might lead to reduced, rather than increased, efficiency (Figlio and O'Sullivan 2001). It is true that, in the case of increased accountability, the incentives are less one-sided. Specifically, even if the rent-seeking administrator model is a correct representation of school decision-making, this model is consistent with increased resources and attention being paid to factors that might improve student outcomes in an atmosphere of increased accountability. On the other hand, the same types of models would suggest that school administrators might substitute resources away from productive uses not covered under the accountability system to improve performance in the areas specifically being considered. The evidence from tax and expenditure limitations, therefore, implies that increased accountability may not lead to increased efficiency. Accountability policies should be structured with this lesson in mind.

### DOES PUBLIC SECTOR COMPETITION RAISE ALL BOATS? IMPACT OF CHARTER SCHOOLS ON STUDENT PERFORMANCE

For a segment of the education market that serves such a small fraction of students (about 1 percent nationally in 2000–01), charter schools have received a seemingly inordinate amount of attention in the popular press and in general discussion of education reform. The centrality of charter schools in the popular discussion of education reform is signaled by the fact that public school choice and increased federal support for charter schools were two of the major provisions of the No Child Left Behind Act of 2001 that was approved by Congress and signed by President Bush in January 2002. Across the political spectrum, policymakers appear to accept the argument of proponents of charter schools—they "can strengthen public education by promoting competition and liberating innovators from the shackles of tradition" (Toch 1998, p. 34).

Whether charter schools will, in fact, fulfill this promise remains uncertain. What is certain is that the character of a state's charter schools depends critically on how state policymakers spell out the details of charter school financing (see Gill et al. 2001 for further discussion). The financing decisions state policymakers must wrestle with include how much money follows each pupil who enrolls in a charter school, whether start-up funds will be available for charter schools, and whether state moneys will be made available to assist charter schools in securing facilities. Even decisions about whether to allow existing private schools to convert to charter status have significant financial implications. Charter schools, therefore, have the potential to necessitate major changes in a state's system of school finance. And charter schools certainly alter the education landscape, as public school officials in Mesa, Arizona (Toch 1998) and Inkster, Michigan (Wildavsky 1999) have seen.

The relative newness of the charter school movement<sup>11</sup> has meant that the research on the impact of this movement is in the formative stage. In most states, charter schools are simply too new and too small a part of the education sector for any measurable effect to be expected.<sup>12</sup> There are, however, a few states in which the charter school sector has begun to mature. Several authors have taken advantage of this maturation to quantify the effects of the entry of charter schools.<sup>13</sup>

In the earliest of these studies, Bettinger (1999) uses data from Michigan to address one of the central questions in the public school choice debate: Will the presence of charter schools (or of other choices in the public sector) improve the performance of public school students who do and do not attend choice schools? The available data allowed him to examine school-level performance measures, control for student performance at the time the first cohort of test-takers entered the charter school, and account for a rich set of student demographic characteristics. In his preferred specifications, Bettinger finds little evidence of improvement in charter schools in the test performance of successive cohorts of fourth and seventh graders. In fact, some relative decline in performance is apparent in his estimates. Further, even after accounting for the possible endogeneity of charter school location, Bettinger observes no relationship between student performance in traditional public schools and the extent of charter school entry. These results would appear to support the conclusion that charter schools fail to generate direct or indirect improvement in student performance.

For several reasons, however, the Bettinger results cannot be viewed as the final word on charter schools. First, as Bettinger himself notes, he is unable to quantify the long-run effects of charter schools. Further, he notes that the poor performance of the charter schools in his sample may be attributable to "institutional immaturity" (p. 21). Also, since the charter schools in Bettinger's data are relatively new, many of the students being tested will be finishing their first year in a new school

<sup>&</sup>lt;sup>11</sup> The first legislation permitting the creation of charter schools was enacted in Minnesota in 1991. And much of the growth of charter schools has occurred over the last several years, with the number of children in charter schools tripling over the last three academic years. In addition, in 1999–00 only in seven states were more than 1 percent of all students enrolled in charter schools, with the District of Columbia, Arizona, and Michigan exhibiting the most entry.

<sup>&</sup>lt;sup>12</sup> The relative smallness of the charter school sector would not preclude estimating the effect of charter schools on students attending those schools. But, if charter schools are not seen by traditional public schools as being real competitors, it is unlikely that any competitive effects will be observed.

<sup>&</sup>lt;sup>13</sup> For far more thorough reviews of the research on charter schools and their effects, see Gill et al. (2001) and Miron and Nelson (2001).

environment. Performance declines would be expected for such students, as such declines after changing schools are well documented in the literature (O'Brien 2002). In addition, Bettinger cannot track cohorts and, therefore, cannot control adequately for pre-charter performance of the cohorts who are tested in the first and second years of charter school operation. Finally, since charter school policies vary dramatically from state to state, lessons from one state may not apply to others. Still, Bettinger's results are hardly a ringing endorsement for charter schools.

Three more recent studies support the argument that Bettinger's results on the impact of charter schools on their own students may understate the long-run impact. In the first of these studies, Eberts and Hollenbeck (2001) examine individual-level test score data on fourth and fifth graders in traditional and charter public schools in Michigan. To create their comparison group of students in traditional public schools, Eberts and Hollenbeck determine the public school district in which each charter school in Michigan was located and include in their sample all of the students in traditional public schools in that district. Like Bettinger, they find some evidence of lower levels of and smaller gains in test scores for students in charter schools.<sup>14</sup> However, when Eberts and Hollenbeck control for the length of time for which a charter school had been opened, they find that the gaps between the performance of students in traditional and charter schools are smaller the longer the charter school had been in operation. Bettinger's suggestion that "institutional immaturity" matters appears to be correct.

It also appears that the performance of individual students increases as those students spend more time in the charter school. Eberts and Hollenbeck do not examine this possibility, in part because they lack the yearly test score data that make examining gains feasible. Two studies, one for Texas (Gronberg and Jansen 2001) and one for Arizona (Solmon, Paark, and Garcia 2001), are able to consider gains and, therefore, can isolate the effect of time spent in the charter school. In both studies, the test scores of students decline in their first year in a charter school. However, Solmon, Paark, and Garcia find that as students spend more time in charter schools their test scores rise relative to their counterparts in traditional public schools.<sup>15</sup> For students in charter schools in which a

<sup>&</sup>lt;sup>14</sup> Actually, Eberts and Hollenbeck cannot control for previous test performance in the same subject. In the first of their gains equations, they use a student's fourth grade test score in mathematics as a control when estimating the impact of charter school attendance on fifth grade science test scores. Similarly, they use the student's fourth grade reading score as the pre-test score when examining fifth grade writing test scores.

<sup>&</sup>lt;sup>15</sup> Nelson and Hollenbeck (2001) raise a number of methodological concerns about the Solmon, Paark, and Garcia analysis. Nelson and Hollenbeck's principal suggestion is that, given the nature of the Arizona data, the evaluation of the impact of charters should be limited to only those students who were in their first year in a charter school. For the reasons noted above, estimating the impact of charters using only recent movers is likely to

disproportionately high share of the students are at-risk, Gronberg and Jansen observe a similar relative increase.

These latter two studies do not, however, contradict all of Bettinger's findings. Gronberg and Jansen find relative performance declines in those charter schools serving disproportionately low shares of at-risk students. Further, Gronberg and Jansen's estimates indicate that student performance is particularly low in start-up charter schools.

As this brief review indicates, the evidence on the impact of charter schools on student performance is decidedly sparse. New charter schools need time to become established; relative performance in these schools is likely to be low in their first and even their second year. Whether, in the long run, charter schools in some states raise the test scores of their students remains an open question, particularly given the tremendous variation across states in the structure of charter school programs and the differences between the Solmon, Paark, and Garcia and Gronberg and Jansen studies in the estimates of the long-run impact of charter schools on the performance of non-at-risk students.

The long-run impact of charter schools on students who remain in traditional public schools also remains an open question. As was noted above, Bettinger finds that charter school entry results in no significant change in the relative performance of students who remain in traditional public schools located in the drawing area of the charter school. Eberts and Hollenbeck present a mixed picture of the competitive effects of charter schools. In their preferred specification, they find that fifth grade science and writing scores are relatively higher in those traditional public schools situated in districts in which charter schools are located. However, fourth grade math scores are relatively lower in such schools, and fourth grade reading scores are not significantly different.

Like Bettinger and Eberts and Hollenbeck, Hoxby (2001b) attempts to estimate the competitive impact of charter schools in Michigan. She also examines the competitive impact of charter schools in Arizona, the other state in which the charter school sector is relatively mature. Hoxby argues that only in districts in which charter schools serve at least 6 percent of the students would we expect to see noticeable competitive effects. Thus, her empirical strategy is to ask whether, in those traditional public schools in districts where charter schools serve at least 6 percent, the growth in student achievement has been faster than in those districts in which the 6 percent threshold has not been crossed. In both Michigan and

understate significantly the long-run impact of charters. Thus, this particular methodological concern seems misguided. The remaining concerns of Nelson and Hollenbeck have considerable merit; whether accounting for these concerns would overturn the results of Solmon, Paark, and Garcia remains an open question.

Arizona, she finds that, particularly at the fourth grade level, the growth has been faster in those districts with substantial charter entry.

The results of Eberts and Hollenbeck's and Hoxby's (2001b) studies would appear to imply that, in fact, charters may well generate a positive competitive effect. And, the fact that both of these studies generate stronger competitive effects than Bettinger could be explained by the fact that Eberts and Hollenbeck and Hoxby are examining public school systems in which the charter school sector is mature and in which traditional public schools have had the opportunity to respond to their new competitors. But, the differences between these latter two studies and that of Bettinger could also be attributable to critical methodological differences. For example, Bettinger correctly observes that controlling for the endogeneity of charter school location is critical. Otherwise, the possibility exists that improvement in the traditional public schools is driven not by charter school entry but by some unobservable factor that drives both charter entry and test score gains in the traditional public schools.<sup>16</sup>

Since neither Eberts and Hollenbeck nor Hoxby account for endogeneity, their estimates of competitive effects must be treated with caution. Similarly, only Bettinger is able to include compelling controls for the pre-test status of students. In other words, when they estimate competitive effects neither Eberts and Hollenbeck nor Hoxby completely rule out the possibility that differences in the cohorts of students tested drive the estimated effects. Finally, while Hoxby's argument that we would expect to see competitive responses only in those districts in which the charter school presence is sufficiently large is compelling, her choice of a 6 percent threshold seems arbitrary. Further, she gives no indication how the results would change if that threshold were lowered. The reality is that the results of Eberts and Hollenbeck and Hoxby do not provide definitive estimates of the competitive effects of charter schools.

What is apparent is that none of the extant research supports the conclusion that the charter school movement will do irreversible damage to the students served by charter schools or to those who remain in traditional public schools. Even the worst-case estimates indicate that relative performance declines in charter schools are small and that students who remain in traditional public schools are essentially unaffected. And, even if the small declines in the performance of charter school students are real, these declines must be balanced against the increased satisfaction of parents of children in charter schools (Gill et al. 2001).

<sup>&</sup>lt;sup>16</sup> Betts (2002) gives one example of such a factor.

### **CONCLUDING REMARKS**

The preceding review of three major changes in the school finance landscape indicates that, while we have learned much from previous research, much still needs to be learned about the effects of these changes. As is apparent from the most recent charter school studies, new data sets in which students are tracked over time will make it easier for researchers to quantify conclusively the effects of policy changes. What may be less apparent from the preceding discussion is the need for researchers to acknowledge that policies that have the same name in two states may actually be very different. School finance reforms, tax and expenditure limitations, and legislation enabling the creation of charter schools have as many differences across states as they have commonalities. The challenge facing researchers is to determine what lessons can be learned only from national-level analyses and only from state-level case studies and to distill these lessons for policymakers. The recent review by Gill et al. on the evidence of choice is a nice example of the type of work that will need to be an essential part of future research.

#### References

- Aaronson, Daniel. 1999. "The Effect of School Finance Reform on Population Heterogeneity." National Tax Journal 52 (1): 5-29.
- Bénabou, Roland. 1996. "Equity and Efficiency in Human Capital Investment: The Local Connection." Review of Economic Studies 63 (2): 237-64.
- Bettinger, Eric. 1999. "The Effect of Charter Schools on Charter Students and Public Schools." National Center for the Study of Privatization in Education Occasional Paper No. 4, Columbia University.
- Betts, Julian R. 2002. "Discussion: Do State Governments Matter?" this volume.
  Bishop, John. 1989. "Is the Test Score Decline Responsible for the Productivity Growth Decline?" American Economic Review 79 (1): 178–197.
  - -. 1992. "The Impact of Academic Competencies on Wages, Unemployment, and Job Performance." Carnegie-Rochester Conference Series on Public Policy 37 (December): 127-94.
- Bradbury, Katharine L., Karl E. Case, and Christopher J. Mayer. 1998. "School Quality and Massachusetts Enrollment Shifts in the Context of Tax Limitations." New England Economic Review (July/August): 3-20.
- Brunner, Eric and Jon Sonstelie. 1997. "Coping With Serrano: Voluntary Contributions to California's Public Schools." In Proceedings of the 89th (1996) Annual Conference on Taxation. Washington, DC: National Tax Association.
- Campaign for Fiscal Equity. 2001. In Evidence: Policy Reports from the CFE Trial. Vol 3. New York: Campaign for Fiscal Equity.
- Card, David and A. Abigail Payne. 2002. "School Finance Reform, the Distribution of School Spending, and the Distribution of Student Test Scores." Journal of Public Economics 83  $(\bar{1}): 49 - \bar{82}.$
- Courant, Paul N. and Susanna Loeb. 1997. "Centralization of School Finance in Michigan." Journal of Policy Analysis and Management 16 (1): 114-36.
- Cutler, David M., Douglas W. Elmendorf, and Richard J. Zeckhauser. 1999. "Restraining the Leviathan: Property Tax Limitations in Massachusetts." Journal of Public Economics 71 (3): 313-34.
- Dee, Thomas S. 2000. "The Capitalization of Education Finance Reforms." Journal of Law and Economics 43 (1): 185-214.

Downes, Thomas A. 1992. "Evaluating the Impact of School Finance Reform on the Provision of Public Education: The California Case." *National Tax Journal* 45 (4): 405–19.
 ——. 1996. "An Examination of the Structure of Governance in California School Districts Before and After Proposition 13." *Public Choice* 86 (March/April): 279–307.

——. 2002. "School Finance Reform and School Quality: Lessons from Vermont." Tufts University, working paper.

- Downes, Thomas A. and David N. Figlio. 1999a. "Do Tax and Expenditure Limits Provide a Free Lunch? Evidence on the Link Between Limits and Public Sector Service Quality." *National Tax Journal* 52 (1): 113–28.
- ——. 1999b. "What Are the Effects of School Finance Reforms? Estimates of the Impact of Equalization on Students and Affected Communities." Tufts University, working paper.
- ——. 2000. "School Finance Reforms, Tax Limits, and Student Performance: Do Reforms Level Up or Dumb Down?" Tufts University, working paper.
- ———. 2001. "Tax Revolts and School Performance." In *Improving Educational Productivity*, edited by D. H. Monk, H. J. Walberg, and M. C. Wang. Greenwich, CT: Information Age Publishing.
- Downes, Thomas A. and Shane M. Greenstein. 1996. "Understanding the Supply of Non-Profits: Modeling the Location of Private Schools." *RAND Journal of Economics* 27 (2): 365–90.
- Downes, Thomas A. and David Schoeman. 1998. "School Finance Reform and Private School Enrollment: Evidence from California." *Journal of Urban Economics* 43 (3): 418–43.
- Downes, Thomas A. and Mona Shah. 1995. "The Effect of School Finance Reform on the Level and Growth of Per Pupil Expenditures." Tufts University Working Paper No. 95-4.
- Downes, Thomas A., Richard F. Dye, and Therese J. McGuire. 1998. "Do Limits Matter? Evidence on the Effects of Tax Limitations on Student Performance." *Journal of Urban Economics* 43 (3): 401–17.
- Duncombe, William and Jocelyn M. Johnston. 2002. "Is Something Better Than Nothing? An Assessment of School Finance Reform in Kansas." Center for Policy Research, Maxwell School of Citizenship and Public Affairs, Syracuse University, working paper.
- Eberts, Randall W. and Kevin M. Hollenbeck. 2001. "An Examination of Student Achievement in Michigan Charter Schools." Upjohn Institute Staff Working Paper No. 01-68.
- Epple, Dennis and Richard Romano. 1998. "Competition Between Public and Private Schools, Vouchers, and Peer-Group Effects." American Economic Review 88 (1): 33–62.
- Epple, Dennis, David Figlio, and Richard Romano. 1998. "Stratification and Peer Effects in Education: Evidence Using Within-school and Between-school Variation in the Data." University of Florida, working paper.
- Evans, William N., Sheila Murray, and Robert M. Schwab. 1997. "Schoolhouses, Courthouses, and Statehouses after *Serrano." Journal of Policy Analysis and Management* 16 (1): 10–31.

—. 1999. "The Impact of Court-Mandated School Finance Reform." In *Equity and Adequacy in Education Finance: Issues and Perspectives,* edited by H.F. Ladd, R. Chalk, and J.S. Hansen. Washington, DC: The National Academies Press.

Fernandez, Raquel and Richard Rogerson. 1997. "Education Finance Reform: A Dynamic Perspective." Journal of Policy Analysis and Management 16 (1): 67–84.

——. 1998. "Public Education and Income Distribution: A Dynamic Qualitative Evaluation of Education Finance Reform." *American Economic Review* 88 (4): 813–33.

- Figlio, David N. 1997. "Did the 'Tax Revolt' Reduce School Performance?" *Journal of Public Economics* 65 (3): 245–69.
- Figlio, David N. and Arthur O'Sullivan. 2001. "The Local Response to Tax Limitation Measures: Do Local Governments Manipulate Voters to Increase Revenues?" *Journal of Law and Economics* 44 (1): 233–58.
- Fischel, William A. 1989. "Did Serrano Cause Proposition 13?" National Tax Journal 42 (4): 465–73.

——. 1996. "How Serrano Caused Proposition 13." Journal of Law and Politics 12 (4): 607–45. Fisher, Ronald C. 1996. State and Local Public Finance, 2<sup>nd</sup> ed. Chicago: Richard D. Irwin, Inc.

- Flanagan, Ann and Sheila Murray. 2002. "A Decade of Reform: The Impact of School Reform in Kentucky." RAND Corporation, working paper.
- Friedman, Milton. 1955. "The Role of Government in Education." In *Economics and the Public Interest*, edited by R. A. Solo. Piscataway, NJ: Rutgers University Press.
- Galles, Gary M., and Robert L. Sexton. 1998. "A Tale of Two Jurisdictions: The Surprising Effects of California's Proposition 13 and Massachusetts' Proposition 2<sup>1</sup>/<sub>2</sub>." American Journal of Economics and Sociology 57 (2): 123–33.
- Gill, Brian P., P. Michael Timpane, Karen E. Ross, and Dominic J. Brewer. 2001. *Rhetoric Versus Reality: What We Know and What We Need to Know About Vouchers and Charter Schools*. MR-1118-EDU. Santa Monica, CA: RAND Corporation.
- Greiner, John M. and George E. Peterson. 1986. "Do Budget Reductions Stimulate Public Sector Efficiency? Evidence from Proposition 2½ in Massachusetts." In *Reagan and the Cities*, edited by G. E. Peterson and C. W. Lewis. Washington, DC: Urban Institute Press.
- Gronberg, Timothy J. and Dennis W. Jansen. 2001. Navigating Newly Chartered Waters: An Analysis of Texas Charter School Performance. Austin, TX: Texas Public Policy Foundation.
- Hanushek, Eric A. 1986. "The Economics of Schooling: Production and Efficiency in the Public Schools." Journal of Economic Literature 24 (3): 1141–77.
- ———. 1996. "School Resources and Student Performance." In Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success, edited by G. Burtless. Washington, DC: Brookings Institution.
- Hanushek, Ēric A. and Julie Ā. Somers. 1999. "Schooling, Inequality, and the Impact of Government." NBER Working Paper No. 7450 (December).
- Hoxby, Caroline M. 2001a. "All School Finance Equalizations Are Not Created Equal." Quarterly Journal of Economics 116 (4): 1189–1231.
- 2001b. "How School Choice Affects the Achievement of *Public* School Students." Harvard University, working paper.
- Husted, Thomas A. and Lawrence W. Kenny. 2000. "Evidence on the Impact of State Government on Primary and Secondary Education and the Equity-Efficiency Tradeoff." Journal of Law and Economics 43 (1): 285–308.
- Joint Budget Committee. 1979. An Analysis of the Effects of Proposition 13 on Local Governments. Sacramento, CA: California Legislature.
- Kirchgässner, Gebhard. 2001. "The Effects of Fiscal Institutions on Public Finance: A Survey of the Empirical Evidence." CESifo Working Paper No. 617.
- Manwaring, Robert L. and Steven M. Sheffrin. 1997. "Litigation, School Finance Reform, and Aggregate Educational Spending." *International Tax and Public Finance* 4 (2): 107–27.
- Mathis, William J. 1998. "Act 60 and Proposition 13." Montpelier, VT: Concerned Vermonters for Equal Educational Opportunity. <a href="http://www.act60works.org/oped1.html">http://www.act60works.org/oped1.html</a> 24 July 2002.
- McClaughry, John. December 1997. "Educational Financing Lessons from California." Concord, VT: Ethan Allen Institute. <a href="http://www.ethanallen.org/commentaries/1997/educatingfinancial.html">http://www.ethanallen.org/commentaries/1997/educatingfinancial.html</a> 24 July 2002.
- McGuire, Therese J. 1999. "Proposition 13 and Its Offspring: For Good or for Evil?" National Tax Journal 52 (1): 129–38.
- Miron, Gary and Christopher Nelson. 2001. "Student Academic Achievement in Charter Schools: What We Know and Why We Know So Little." National Center for the Study of Privatization in Education Occasional Paper No. 41, Columbia University.
- Murnane, Richard J., John B. Willett, and Frank Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." *Review of Economics and Statistics* 77 (2): 251–66.
- Murray, Sheila E. 2001. "State Aid and Education Outcomes." In *Improving Educational Productivity*, edited by D. H. Monk, H. J. Walberg, and Margaret C. Wang. Greenwich, CT: Information Age Publishing.
- Murray, Sheila E., William N. Evans, and Robert M. Schwab. 1998. "Education Finance Reform and the Distribution of Education Resources." *American Economic Review* 88 (4): 789–812.
- Neal, Derek and William Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104 (5): 869–95.
- Nechyba, Thomas J. 1996. "Public School Finance in a General Equilibrium Tiebout World:

Equalization Programs, Peer Effects and Private School Vouchers." NBER Working Paper No. 5642 (June).

———. 2000. "Mobility, Targeting, and Private School Vouchers." American Economic Review 90 (1): 130–46.

- Nelson, Christopher and Kevin Hollenbeck. 2001. "Does Charter School Attendance Improve Test Scores? Comments and Reactions on the Arizona Achievement Study." W.E. Upjohn Institute Staff Working Paper No. 01-70.
- O'Brien, Daniel M. 2002. "The Impacts of Structural Mobility on Student Achievement." School of Social Sciences, University of Texas at Dallas, working paper.
- O'Sullivan, Arthur, Terri A. Sexton, and Steven M. Sheffrin. 1995. Property Taxes and Tax Revolts: The Legacy of Proposition 13. Cambridge, England: Cambridge University Press.
- Schwadron, Terry, editor. 1984. California and the American Tax Revolt: Proposition 13 Five Years Later. Berkeley: University of California Press.
- Shadbegian, Ronald J. 2001. "Did Proposition 2<sup>1</sup>/<sub>2</sub> Affect Local Public Education in Massachusetts? Evidence from Panel Data." University of Massachusetts, Dartmouth, working paper.
- Silva, Fabio and Jon Sonstelie. 1995. "Did Serrano Cause a Decline in School Spending?" National Tax Journal 48 (2): 199–215.
- Solmon, Lewis, Kern Paark, and David Garcia. 2001. *Does Charter School Attendance Improve Test Scores? The Arizona Results.* Phoenix: The Goldwater Institute.
- Stocker, Frederick D. 1991. "Introduction." In Proposition 13: A Ten-Year Retrospective, edited by F. D. Stocker. Cambridge, MA: Lincoln Institute of Land Policy.
- Tiebout, Charles M. 1956. "A Pure Theory of Local Public Expenditures." *Journal of Political Economy* 64 (5): 416–24.

Toch, Thomas. 1998. "The New Education Bazaar." U.S. News and World Report (27 April).

- Vigdor, Jacob L. 1998. "Local Taxes and the Growth of Cities." Harvard University, working paper.
- Wildavsky, Ben. 1999. "Why Charter Schools Make Inkster Nervous." U.S. News and World Report (28 June).
- Wolfe, Barbara and Robert Haveman. 2002. "Social and Nonmarket Benefits from Education in an Advanced Economy," this volume.

## Discussion

## **Do State Governments Matter?**

## Julian R. Betts\*

Tom Downes is ambitious in his goal to summarize the impact of three distinct types of state policy changes on the ability of school systems to "equalize opportunity" across students of varying socioeconomic backgrounds. The policy reforms he considers are finance reforms emanating from court challenges to individual states' school finance systems, reforms deriving from state tax and expenditure limitations, and the advent of charter schools as a publicly funded alternative to regular public schools.

To those not familiar with the debate on whether educational spending matters, it is worth mentioning that the works that Downes reviews on court decisions and voter tax limitations are of great interest to economists studying public education. One of the most important areas in education economics is the extent to which changes in school funding cause changes in outcomes such as test scores, graduation rates, college attendance, and earnings of students years after graduation. A large body of literature examines the relationship between school resources and student achievement, years of schooling completed, and earnings after leaving school.<sup>1</sup> But does a positive correlation between school expenditures and student outcomes necessarily imply causation? There are many reasons to think not. Most obviously, in the United States today, students of lower socioeconomic background still typically attend schools with lower levels of resources, particularly when resources are measured by teacher qualifications such as credentials, years of experi-

<sup>\*</sup> Professor of Economics, University of California, San Diego.

<sup>&</sup>lt;sup>1</sup> For a review of the test-score literature, see Hanushek (1996); for a review of the relation between school resources and years of schooling and earnings, see Betts (1996).

ence, and education. If researchers find that disadvantaged students have both poorer educational outcomes and fewer resources at school, it certainly could signal that resources do matter. But the correlation could equally well be spurious. For instance, it could be that the true reason that disadvantaged students tend to have poorer educational outcomes is that they receive fewer educational resources in the home, fewer supports among the family, and fewer highly educated role models in the local community. In this instance, the positive correlation between school resources and student outcomes is merely that: a correlation induced by imperfectly measured variations in student socioeconomic status.

Of course, economists do more than look at simple correlations. Regression analysis attempts to control for all factors that may affect the dependent variable, in this case student outcomes. But we lack data sources that include rich measures of the educational supports in the home and community that I listed above. It is unlikely that commonly available measures of socioeconomic status, such as parental education and eligibility for school lunch assistance, fully capture variations in these factors. Hence, even careful regression analysis might overstate the impact of school resources on student outcomes because both variables are positively correlated with imperfectly observed family and community resources.

Conversely, one could argue that the many attempts by federal, state, and local governments to provide compensatory educational aid to schools in impoverished neighborhoods could induce a negative correlation between school resources and socioeconomic status. This, in turn, could induce a negative correlation between student outcomes and school resources that again is not causal, but merely reflects the correlations between both these variables and student disadvantage.

On the whole, I find the first of these arguments more persuasive, as most of our rather imperfect measures of student socioeconomic status tell us that there is still a predominantly positive relationship between socioeconomic status and the level of school resources that a student receives. In addition, few can doubt that socioeconomic status is a powerful determinant of cognitive development in children. For instance, a recent Educational Testing Service (ETS) study (Coley 2002) finds extremely large gaps in various measures of academic achievement between students with low and high socioeconomic status at the *start* of kindergarten. This surely speaks to the major contributions of home and neighborhood to early cognitive development.

# COURT-MANDATED EDUCATIONAL REFORMS AND TAX AND EXPENDITURE LIMITATIONS

I would now like to discuss the court-mandated educational reforms and the tax and expenditure limitations that Downes reviews. Both of


these literatures attempt to reduce problems of endogenous school spending and omitted variable bias by seeking exogenous sources of variation in the resources that a school receives. In the case of court cases, one can argue that both the launching of a school finance lawsuit and especially the timing of its resolution are not caused by unobservable demographic or other personal attributes of state residents, or by any other characteristics of the state that could be causally related to student outcomes and school resources. If this assumption holds, then social scientists can perform before-and-after analyses of student outcomes that can potentially provide unbiased estimates of the impact of changes in school finances on student performance. In the more sophisticated approach that has quickly come to the fore, economists instead perform "difference-in-difference" analyses that compare changes in student outcomes over time in states that have undergone court-mandated finance reforms to changes in states that have not been subject to court mandates. This approach effectively takes account of national trends in the underlying variables and unobserved and constant characteristics of each state.

To those not familiar with difference-in-difference models, a simple example may be helpful. Figure 1 shows average annual gains in students' test scores plotted against spending per pupil in two hypothetical states, for two different years. The state in the upper left of the figure habitually spends less on schools but has higher rates of student learning, perhaps because of some other unobserved factors affecting both variables. (In this hypothetical world, perhaps fiscally conservative parents not only vote to spend less on schools but also read more to their children at home!)

Let us suppose that court decisions in both states have caused spending per pupil to rise slightly, which in turn has quite literally *caused* student learning rates to increase in both state A and state B, as shown. But linear regression would not detect these causal effects. As shown by the dotted line representing the fitted regression line in Figure 1, we obtain the "wrong" result. There appears to be a negative relation between spending per pupil and student learning because the betweenstate differences in spending and learning completely dominate the visible but small effects of increases in spending per pupil in each state.

The difference-in-difference estimation strategy solves this problem by comparing changes in one state to changes in another. Social scientists typically estimate these models by expanding the list of explanatory variables from spending per pupil alone to also include a set of dummy (0,1) variables for states.<sup>2</sup> It can be shown that this is equivalent to subtracting the state mean from both gains in achievement and spending per pupil from each observation, and then running a linear regression using these "de-meaned" variables. Figure 2 illustrates what happens when we subtract the state means in this way. The changes over time in states A and B now line up perfectly along a positively sloped line. When we estimate a linear regression on these transformed data, we correctly estimate a positive causal relation between spending and learning. The trick in this analysis is to disregard all of the between-state variation, and instead focus only on the within-state variation.

The tax-limitation and expenditure-limitation literature works on a similar premise: If voters pass these limitations for reasons that are not related to student outcomes in the state, then economists often consider the resulting reduction in school spending as occurring exogenously with respect to student outcomes. What we have, in both cases, is a natural experiment in which some outside or exogenous force has induced a change in school finance.

Downes reviews these literatures with care. He correctly concludes that the existing literature on court-mandated school finance changes has yet to deliver a consistent message about the impact on either the level or the distribution of student outcomes. The tax-limit and spending-limit literature provides slightly more definitive results suggesting that mean performance may fall if spending per pupil drops because of limitations.

Despite these methodological advances, using state-level variation in

<sup>&</sup>lt;sup>2</sup> To keep the analysis simple, for this example, I will ignore the additional control or controls for time trends that researchers typically employ.





court decisions or tax limits carries certain risks. The crucial assumptions here are that court rulings on education finance and voter passage of tax or expenditure limitations occur in ways that are exogenous with respect to student outcomes. One can imagine scenarios in which either type of event occurred endogenously with respect to school quality. For instance, suppose that lower-income parents in one state become increasingly concerned about the quality of public schooling. This increased concern could manifest itself in several ways, for instance, in increased parental involvement in schools, which might improve student outcomes in these less-affluent areas. At the same time increased parental concern could lead to a lawsuit seeking to equalize school spending between have and have-not districts. If the court case is successful, these two events would lead—separately—to an increase in test scores in disadvantaged districts and an increase in school spending in the very same districts. Although a difference-in-difference analysis would lead us to infer that increased school spending had improved test scores, in reality, both changes would have been caused by something quite different-increased parental activism in the have-not districts.

A weakness in the above argument is that it ignores the fact that legal challenges to states' systems of education finance can typically take years and in some cases decades to draw to a final conclusion. This would make the timing of the increase in test scores and the court-ordered change in spending less coincident.

Another example, this one related to the passage of tax or expenditure limitations, is that voters are more likely to support such limitations if they come to believe that state and local governments are not spending current tax revenues effectively. One event that could spur such a belief among voters is a downward trend (or stagnation) in student achievement, in spite of recent increases in spending per pupil. (Such increases in spending have been the norm over the last half-century.) This leads us into a situation of reverse causation, in which a decline or stagnation in student achievement causes the tax limitation measure to pass. It is not hard to see how even a very careful researcher might misconstrue this correlation as meaning that the new tax limitation had caused test scores to decline. Only by carefully removing ongoing trends in both variables can the researcher hope to obtain the correct inference.<sup>3</sup> The underlying issue in this second example is that the primary identification approach used in both literatures, the difference-in-difference method, is prone to error because before-and-after analyses can mistakenly attribute differing trends in different states to the change in policy. My goal here is not to dismiss the literatures that exploit the apparent exogeneity of court orders and tax limitations. On the contrary, they represent important developments in the broader literature on the determinants of school quality. Rather, my goal is to caution that the research and education policy communities would be wrong to treat either approach as a panacea.

Downes provides a careful and evenhanded summary of the findings that emerge from the court-mandate and the tax- and expenditurelimitation literatures. The results vary across data sets and the specific techniques used, which in part may reflect occasional violations of the assumptions underlying the natural experiments that these papers study. Overall, the body of work summarized by Downes suggests that changes in school spending are related to student outcomes in the expected direction, although the court-mandate literature is murkier in this regard than the tax- and expenditure-limitation work. My own reading of these papers is that the effects are modest, in the sense that complete equalization of school funding would go only part way toward equalizing student achievement.

The related literature on the impact of school resources on the earnings of students in the years after graduation points in the same direction. Betts and Roemer (2001) use the National Longitudinal Survey of Young Men to address the question of the extent to which educational

<sup>&</sup>lt;sup>3</sup> Downes and Figlio (2000) represent a good attempt to tackle this specific possibility head on, and more work of this nature needs to be done.

funds would have to be reallocated across students to equalize opportunity across groups, defined, following Roemer (1998), as equalizing wages in an average sense. We find that equalizing spending per pupil, for instance between black and white students, would do virtually nothing to equalize the black–white wage gap years after the students had left school. Rather, spending per pupil would have to be many times larger for black students if policymakers wanted to take a significant chunk out of the black–white wage gap.

My final observation on Downes's summaries of the court-mandate and the tax- and expenditure-limitation literatures is that, as a central contributor to these literatures, he has written an extremely balanced review that points out the limitations not only of others' work but of his own. This is a model for others to follow.

## CHARTER SCHOOLS

The third avenue of research reviewed by Downes is the advent of charter schools as an alternative to regular public schools. He asks whether students attending charter schools increase their rate of learning once enrolled, and the more difficult question of whether the advent of charter schools as a competitive force has induced regular public schools to improve.

Downes discusses two recent evaluations of charter schools in Texas and Arizona. These evaluations suggest a first-year slump for students enrolled in charter schools, followed by improvements for at least some charter school students in later years. This finding is of great importance given that school districts typically place charter schools under the accountability microscope practically from day one of their establishment. It will be important to see whether these dynamics can be replicated in other states. If so, administrators should be apprised of these patterns in order to avoid over-reacting to initial results at startup charter schools. At present, the results are not sufficiently solid for us to know for sure. (For a critique of the Arizona study, see Nelson and Hollenbeck 2001.)

On the question of whether the establishment of charter schools creates competitive pressures that spur nearby regular public schools to improve, Downes discusses the Michigan work of Bettinger (1999) at some length. Again, his review is on target in that data limitations restrict what we can know with certainty. The tentative conclusion from Bettinger's work is that we lack evidence of competitive pressures that improve student achievement at public schools located near charter schools.

However, another study, Hoxby (2002), finds statistically significant evidence that test scores, relative to spending per pupil, rise significantly in districts in which charter schools come to represent 6 percent or more of student enrollment. Hoxby uses a difference-in-difference approach, as do many authors in the two aforementioned literatures. As I argued earlier, such approaches are susceptible to error if there are differences in the trends in student achievement among schools that are, by happenstance, correlated with the enrollment share of charter schools in the local district. To her credit, Hoxby successfully replicates her results by testing for a change in the *trend in gains* in school productivity after charter schools become a significant competitive force.

A key problem that remains, however, is that we do not know why charter schools become commonplace in some districts, yet remain so rare in other districts. It is quite easy to think of circumstances that would bias the estimated impact of charter schools on regular public schools up or down. For instance, suppose that one of the many omitted variables in existing analyses is the quality of district leadership and its openness to change. Suppose that a district hires a new, reform-minded superintendent, who simultaneously implements meaningful reforms in the public schools and, as part of the package, increases the number of charter schools. Even if charter schools had no real impact on the quality of regular public schools, a positive correlation between the number of charter schools in the district and public school productivity would result, and this, again, would not be causal. Even Hoxby's useful de-trended difference-in-difference approach would not capture the true causal relationships in such an instance.

# SUMMING UP: WHAT WE HAVE LEARNED AND WHAT WE STILL DON'T KNOW

While work on the question of charter schools' impact on regular public schools is still in its infancy, we have already learned important lessons. To date, little evidence supports those who warned that charter schools would be an educational disaster, as Downes points out. But we do not have as much positive to say as proponents of charter schools might like. Hoxby's work provides the strongest evidence to date that charter schools might have a positive competitive effect, even if it occurs after a threshold point has been reached. It also provides an intelligent check on the standard difference-in-difference approach.

To applied economists, the court-mandate and tax-limit and expenditure-limit literatures offer important examples of attempts to find exogenous sources of variation in an explanatory variable (in this case, spending per pupil) with the ultimate goal of unearthing the true causal impact of that variable on the outcome of interest (in this case, student achievement). Much of the literature has adopted the difference-indifference approach, which in essence compares changes in outcomes in states (or districts) that have undergone a policy shock (such as a tax limitation) with changes in states that have not experienced the shock, all the while removing fixed characteristics of each state and common trends that occur in all states equally.

The difference-in-difference approach has proven valuable, but it puts us at some risk of attributing changes in one state to the given policy shock when, in fact, another policy innovation or perhaps a demographic shock, imperfectly measured in the researchers' data, was in truth responsible for the change in student outcomes. A specific example of this is when an omitted variable "causes" both the change in student achievement and the change in policy, where the policy change could be either a court decision, a tax or expenditure limit, or the creation of a charter school. In my opinion, this issue is most severe in the charter school literature, where the existence of stark differences across districts in the rate of creation of charter schools suggests an underlying cause, perhaps related to changes in attitudes of the district administration or of local voters. A second risk is that the standard difference-in-difference approach misinterprets variations in trends across states or districts as being caused by the policy change in certain states. As I have noted, some researchers have started to find approaches that at least partially take these concerns into account.

Apart from his review of charter school studies that typically use a district-level or school-by-school analysis, Downes concentrates on lessons from natural experiments at the state level. Readers of the courtmandate and tax and expenditure literatures should be particularly concerned that—at this high level of aggregation—state fixed effects do not do enough to control for unobserved variations among states that, contrary to the assumptions of difference-in-difference work, are not always fixed. Furthermore, the problem of endogeneity (which reforms occur in which jurisdiction) does not disappear at the state level.<sup>4</sup> For these reasons, it will be important to supplement the state-level literatures on supposedly exogenous policy changes with similar analyses at the district level in order to check for consistency.

With these qualifications in mind, we have learned a great deal from all three bodies of literature, in particular from the tax- and expenditurelimit work. But much remains to be done before we can say with reasonable precision and certainty what the exact impact of spending changes or of the creation of charter schools might be on the quality of regular public schools.

<sup>&</sup>lt;sup>4</sup> For a cautionary tale about the dangers of relying on state-level variation to identify the effects of school resources on students' earnings later in life, see Heckman, Layne-Farrar, and Todd (1996).

#### References

- Bettinger, Eric. 1999. "The Effect of Charter Schools on Charter Students and Public Schools." National Center for the Study of Privatization in Education Occasional Paper No. 4, Columbia University.
- Betts, Julian R. 1996. "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature." In *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by G. Burtless. Washington, DC: Brookings Institution.
- Betts, Julian R. and John E. Roemer. 2001. "Equalizing Opportunity through Educational Finance Reform." University of California, San Diego, Department of Economics, unpublished paper.
- Coley, Richard J. 2002. An Uneven Start: Indicators of Inequality in School Readiness. Princeton, NJ: Educational Testing Service.
- Downes, Thomas A. and David N. Figlio. 2000. "School Finance Reforms, Tax Limits, and Student Performance: Do Reforms Level Up or Dumb Down?" Tufts University, working paper.
- Hanushek, Eric A. 1996. "School Resources and Student Performance." In Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success, edited by G. Burtless. Washington, DC: Brookings Institution.
- Heckman, James, Anne Layne-Farrar, and Petra Todd. 1996. "Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship." In Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success, edited by G. Burtless. Washington, DC: Brookings Institution.
- Hoxby, Caroline. 2002. "School Choice and School Productivity (or, Could School Choice be a Tide that Lifts All Boats?)." NBER Working Paper No. 8873 (April).
- Nelson, Christopher and Kevin Hollenbeck. 2001. "'Does Charter School Attendance Improve Test Scores?' Comments and Reactions on the Arizona Achievement Study." W.E. Upjohn Institute Staff Working Paper No. 01-70.
- Roemer, John E. 1998. Equality of Opportunity. Cambridge MA: Harvard University Press.

## Discussion

# **Do State Governments Matter?**

### Michael A. Rebell\*

One of the glories of the American education system is its unique local governance structure, which places substantial responsibility for educational policy in locally elected school boards. Historically related to this governance structure has been a system for financing public education, which is also rooted in local communities and, by and large, is tied to local systems of property taxation.

At one time, variations in the property wealth of local communities were limited and resulted in only mild disparities in the funding available for schooling purposes in different communities. Population shifts to the suburbs in recent decades have resulted in huge differentials in the property values of various urban, suburban, and rural communities, and have vastly exacerbated school funding inequities. As a result, in the twenty-first century, the American system of educational finance, with its emphasis on local real estate property taxation, creates serious injustices. In most states, the reliance on local property values has resulted in the anomalous reality that students with the greatest educational need have the least amount of educational resources available to them.

The demographic and economic growth of the suburbs has been accompanied by a trend toward increasing suburban domination of state legislatures. This has made it difficult for reformers to achieve legislative solutions to funding inequities. Accordingly, residents of property-poor school districts have tended to seek relief in the courts. Almost 30 years ago, a major challenge to the inequities in Texas's system of school

<sup>\*</sup>Executive Director and Counsel, Campaign for Fiscal Equity, Inc., and Adjunct Professor of Law, Columbia University.

finance was brought before the United States Supreme Court in *Rodriguez* v. *San Antonio Independent School District*.<sup>1</sup> The Supreme Court sympathized with the plight of the largely Chicano plaintiffs, whose property-tax rate was approximately 25 percent greater than their neighbors' in the nearby affluent Anglo district, but whose schools had only half of the resources available for their children's education. However, having determined that education was not a "fundamental interest" under the federal courts could not remedy this problem.

To the surprise of many observers, and in one of the most remarkable chapters in the history of state constitutional law, the state courts have, over the past three decades, entered energetically into the fray after the Supreme Court closed the federal courthouse doors. Since the decision in *Rodriguez*, litigations addressing funding inequities have been filed in 44 of the 50 states, and in some states on multiple occasions. Commentators have described various "waves" of outcomes in these cases, marked by varying trends in the reformers' degree of success. (Thro 1990; Levine 1991). Since 1989, there has been a clear trend toward plaintiff success, with reformers prevailing in approximately two-thirds of the 25 major decisions of states' highest courts (Rebell 2002).

Tom Downes's paper provides an excellent survey of the studies that have been undertaken by economists and social scientists in recent years to try to determine the impact of these litigations in terms of (1) reducing disparities in per capita spending among local school districts, (2) increasing overall educational expenditures within a state, and (3) improving student achievement, especially for the most disadvantaged students. Noting the "tremendous diversity of school finance reforms," Downes cautions that any attempt in a national-level study to classify finance reforms will be "imperfect." Accordingly, he recommends complementing these national-level studies with state-level case studies of concrete reforms.

I would go further. My contention is that the tremendous diversity in facts, legal rights and requirements, political context, and specific holdings of courts in various states makes it impossible to draw meaningful conclusions from national-level studies on the impact of fiscal equity litigation. Pursuit of such studies not only misallocates scholarly resources, but the results of these efforts can seriously mislead the public, the press, and policymakers.

I will illustrate this fundamental point by referring to one of the leading studies in this area, undertaken in 1998 by Murray, Evans, and Schwab, which is quoted in Downes's paper.<sup>2</sup> These authors studied the

<sup>&</sup>lt;sup>1</sup> No. 71-1332 Supreme Court of the United States 411 US 1; 93 S. Ct. 1278; 1973.

<sup>&</sup>lt;sup>2</sup> See also Murray, Evans, and Schwab (1999).

outcome of decisions favorable to plaintiffs in 16 states over the period 1972 to 1992. They conclude: "Successful litigation reduced inequality by raising spending in the poorest districts while leaving spending in the richest districts unchanged, thereby increasing aggregate spending on education" (p. 789).<sup>3</sup>

The first problem with this and other national-level studies is the manner in which Murray, Evans, and Schwab identify the states to include in their sample. For example, among the 16 states they study was Alabama, where a trial court issued an extensive reform decision in 1993. Because of the intricacies of Alabama politics, however, the Alabama Supreme Court rejected the remedial order in 1997, and, just this year, formally overruled the 1993 liability decision.<sup>4</sup> Was it reasonable, therefore, to include the Alabama case as an example of a state in which plaintiffs had prevailed? Can any findings concerning a lack of equalization or lack of impact on student achievement over the past 10 years in Alabama fairly be correlated with the ineffectiveness of judicial intervention when there had not actually been any judicial intervention into the educational system?

Consider also Arizona. There, the State Supreme Court issued a major ruling in 1994 that was concerned solely with capital funding.<sup>5</sup> Almost all of the other decisions analyzed by Murray, Evans, and Schwab focus exclusively on operating expenses. Lumping together cases with such different goals and impacts in one analysis of outcomes is highly questionable.

A related point has to do with the time period encompassed by the analysis. The Murray, Evans, and Schwab study covers a broad time frame, incorporating all cases decided during the 23-year period from 1971 to 1994. Since implementation of court decrees is often a lengthy process, it is likely that there will be a greater impact for cases that were decided at the beginning of the study's time period than for those at its end. On the other hand, changes initiated in the early years may be undone by political developments in later years, as occurred in the State of Washington, where reforms initiated to benefit urban areas ultimately came to hurt them (Cipollone 1998).<sup>6</sup>

<sup>&</sup>lt;sup>3</sup> I deliberately chose this work to criticize and to illustrate my thesis partially because their conclusion that litigations, by-and-large, result in productive reforms is highly congenial to me, as an advocate for fiscal equity reform and a lawyer who is currently representing plaintiffs in a major education adequacy litigation. (See *Campaign for Fiscal Equity v. State of New York*, 655 NE.2d 661 (NY 1995), 719 NYS.2d 475 (NY Sup. Ct. 2001), reversed 2002 WL 1369966 (N.Y.A.D. 1 Dept.), appeal pending, N.Y. Ct. App.). Nevertheless, as a legal scholar and social-policy analyst, I must question the validity of this methodology.

<sup>&</sup>lt;sup>4</sup> Alabama Coalition for Equity v. Siegelman, Index No. 1950030, S.Ct, Ala (May 31, 2002).

<sup>&</sup>lt;sup>5</sup> Roosevelt Elementary School Dist. No. 66 v. Bishop, 877 P.2d 806.

<sup>&</sup>lt;sup>6</sup> Evans, Murray, and Schwab (1997) do utilize a ten-year-after reform variable, and conclude that it does not substantially differ from their overall findings (p. 24), but this

The extent to which the years studied can substantially affect the analysis of the impact of a court case in a particular state is starkly illustrated by Michael Heise's 1995 study of the impact of the Connecticut Supreme Court's 1977 ruling in *Horton* v. *Meskill*.<sup>7</sup> Heise concludes that overall the court decision was "associated with declines in state education funding" (p. 212). At the same time, he notes, however, that there was a marked increase in spending during the period 1984 to 1987. Even though the initial decision was issued in 1977, significant action was not taken by the legislature until, in a 1983 follow-up decision, the Court put the legislature on notice that delays in fully funding the new constitutional scheme would not be tolerated. The fact that there were declines in expenditures during the period of noncompliance is not unexpected. The significant fact is that spending sharply increased after the Court's 1983 follow-up decision.

A related problem in this area is that averages, even if accurate, actually tell us little or nothing. Even if all of the court decisions included in a particular categorization are appropriate and the time period involved somehow is fully inclusive, the fact that on average court decisions in a variety of states do or do not have a particular impact provides little useful information for reformers or analysts. Since we know empirically that some court decisions have great impact and others have little or none, average quantitative results are not meaningful. In a 16-state sample, indicators of positive overall impacts may mean that judicial reforms had very strong impacts in three or four states and minor or negative impacts in a dozen others. Or the converse could be true: Broad positive impacts in many states could be countered by strong negative impacts in a few. Should reformers, therefore, look to the courts for relief? Is an investment in litigation worth the time and expense involved? Conclusions based on the averaging of outcomes provide no useful answers to these questions.

Case studies of outcomes in particular states, on the other hand, do provide meaningful information for answering these questions. They can inform us about the precise impact particular judicial interventions have had over specific periods of time. Reasonable conclusions can be drawn about the success of various legal strategies in such empirical analyses. Advocates and researchers considering the relevance of judicial interventions in another state will then be in a position to consider and compare meaningful specific variables.

In short, well-done case studies of the outcomes of litigations in particular states can provide a rich source of data for these analyses.

summary snapshot does not consider the impact of timing variables in each of the particular states.

<sup>7 376</sup> A.2d 359 (Ct. 1977).

Downes's case studies of the impact of judicial decrees in California and Vermont are prime examples of such studies (Downes 1992, 2002). It may also be analytically useful to undertake a large series of related state case studies ("or caselets") and draw conclusions from a qualitative and quantitative assessment of trends revealed by such analyses.<sup>8</sup>

Arguably, the mere fact that litigation has been filed will have a noticeable effect on policymakers and political outcomes. In fact, one study has specifically found that education expenditures tend to rise in states where plaintiffs have filed complaints, whatever the ultimate outcome of the litigation (Hickrod et al. 1992). That conclusion, however, merely substantiates the basic point that judicial interventions can be assessed meaningfully only from within the context of the educational, political, and economic factors at play in a particular state context. The filing of a court case may galvanize attention and push education finance reform to the top of the political agenda. How the matter is handled once it gets that attention, which legal doctrines and which court-ordered remedies do or do not have a positive impact—and why—are the key questions that well-done state-level case studies, but not broad national-level impact studies, can usefully address.<sup>9</sup>

<sup>&</sup>lt;sup>8</sup> For example, see Rebell and Block's 1982 study, the methodology of which is based on 65 "caselets" and four major case studies that review aspects of judicial intervention in educational policy litigations.

<sup>&</sup>lt;sup>9</sup> The extensive debate in the academic literature and the courts for the past two decades on whether "money matters" also illustrates how meta-analyses on a national level can distort major policy discussions in a counterproductive way. More than a decade ago Eric A. Hanushek reported that an overview analysis of approximately 187 distinct studies of the impact of increased funding on student achievement raises serious questions about whether increases in funding can be correlated with positive outcomes in terms of student achievement (Hanushek 1989, 1991). Hanushek's methodologies and conclusion have been strongly challenged (Hedges, Lane, and Greenwald 1994; Card and Krueger 1996).

In the two dozen or so fiscal equity litigations that have taken place since this question arose, huge "battles of the experts" have taken place, and enormous expenditures of time and resources have been devoted to trying to establish whether, in fact, increased funding can lead to improved student achievement. After the dust settled on the academic debate, most of the judges who have focused on this issue in recent cases have reached a common-sense conclusion that money well spent will make a difference, but money merely thrown at the problem may be wasted. (See, for example: *Hoke County Board of Education* v. *North Carolina*, 95 CVS 1158 (Super. Ct., Wake Co.), "Only a fool would find that money does not matter in education"; *Roosevelt Elementary School District 66* v. *Bishop*, 877 P.2d 806, 822 (Ariz. 1994), C.J. Feldman specially concurring, "Logic and experience also tell us that children have a better opportunity to learn biology or chemistry... if provided with laboratory equipment for experiments and demonstrations.") In short then, the issue is not whether money matters, but how to apply appropriate accountability measures to ensure that money that is allocated for education reform is spent in an effective manner.

#### References

- Card, David and Alan B. Krueger. 1996. "Labor Market Effects of School Quality: Theory and Evidence," in *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, edited by G. Burtless. Washington, DC: Brookings Institution.
- Cipollone, Diane W. 1998. "Defining a 'Basic Education': Equity and Adequacy Litigation in the State of Washington." In *Studies in Judicial Remedies and Public Engagement*, edited by Campaign for Fiscal Equity, Inc., vol. 1. New York: Campaign for Fiscal Equity.
- Downes, Thomas A. 1992. "Évaluating the Impact of School Finance Reform on the Provision of Public Education: The California Case." *National Tax Journal* 45 (4): 405–19.
  2002. "School Finance Reform and School Quality: Lessons from Vermont." Tufts University, working paper.
- Evans, William N., Sheila E. Murray, and Robert M. Schwab. 1997. "Schoolhouses, Courthouses, and Statehouses After Serrano." Journal of Policy Analysis and Management 16 (1): 10–31.
- Hanushek, Eric A. 1989. "The Impact of Differential Expenditures on School Performance." Education Researcher 18 (4): 45–51.

——. 1991. "When School Finance 'Reform' May Not Be Good Policy." Harvard Journal on Legislation 28 (2): 423–56.

- Hedges, Larry V., Richard D. Lane, and Rob Greenwald. 1994. "Does Money Matter? A Meta-Analysis of the Effects of Differential School Inputs on Student Outcomes." *Education Researcher* 23 (3): 5–14.
- Heise, Michael. 1995. "The Effect of Constitutional Litigation on Education Finance: More Preliminary Analyses and Modeling." *Journal of Education Finance* 21 (2): 195–216.
- ——. 2001. "The Effect of Constitutional Litigation on Education Finance: More Preliminary Analyses and Modeling." *Journal of Education Finance* 21 (2): 195–216.
- Hickrod, Alan G., Edward R. Hines, Gregory P. Anthony, and John A. Dively. 1992. "The Effects of Constitutional Litigation on Education Finance: A Preliminary Analysis." *Journal of Education Finance* 18 (2): 180–210.
- Levine, Gail F. 1991. "Meeting the Third Wave: Legislative Approaches to Recent School Finance Rulings." Harvard Journal on Legislation 28 (2): 507-42.
- Murray, Sheila E., William N. Evans, and Robert M. Schwab. 1998. "Education-Finance Reform and the Distribution of Education Resources." *American Economic Review* 88 (4): 789–812.

——. 1999. "The Impact of Court-Mandated Finance Reform." In Equity and Adequacy in Education Finance: Issues and Perspectives, edited by H. Ladd, R. Chalk, and J. Hansen. Washington, DC: National Academy Press.

- Rebell, Michael A. 2002. "Education Adequacy, Democracy, and the Courts." In Achieving High Educational Standards for All, edited by T. Reading, C. Eley, Jr., and C. E. Snow. Washington, DC: National Academy Press.
- Rebell, Michael A. and Arthur R. Block. 1982. Educational Policy Making and the Courts: An Empirical Study of Judicial Activism. Chicago: University of Chicago Press.
- Thro, William E. 1990. "The Third Wave: The Implications of the Montana, Kentucky and Texas Decisions for the Future Public School Finance Reform Litigation." *Journal of Law and Education* 19 (2): 219–50.

## Address

# THE CHALLENGE OF TRANSFORMATION

### Michael Barber\*

I am going to discuss the challenge of transformation and education reforms in England. I want to emphasize the importance of Prime Minister Tony Blair's personal commitment to education. Several years ago, he made a speech in which he said his three priorities were "education, education, and education." He has stuck with that agenda well. I do not pretend to be objective or unbiased about this topic. I am very involved and passionate about it. I am very excited about what we are doing.

I am going to address four questions: How are we doing in England? What has worked? What has not worked? What is next?

### How ARE WE DOING?

One of the things that is very important in answering the first question is that we have been on a reform trajectory since the late 1980s. The Blair government did not reverse some of the fundamental steps of reform that were taken under previous governments. As in Texas and North Carolina, this is a reform agenda that has had both parties in power and, in our case, three different Prime Ministers overseeing stages of it. There have been many mistakes along the way, but the narrative is very clear. The best data for international comparisons are the recent OECD Program for International Student Assessment (PISA) results, published in December 2001. In reading literacy among 15-year-olds, the United Kingdom ranked seventh out of the 32 countries; in mathematical

<sup>\*</sup>Head, Prime Minister's Delivery Unit, and Chief Advisor to the Prime Minister on Delivery, United Kingdom.

literacy, we ranked eighth; and in scientific literacy, we ranked fourth.<sup>1</sup> These results surprised most people in England because for the last 30 years we have thought that everybody in either Germany or France, certainly Switzerland, and sometimes the United States and the Far East, is better at education than we are. But on the basis of these results, we compare reasonably well, although when we look at our own system, we are all too aware that for many children it is not yet good enough.

One of the key indicators of our progress over the last few years is in the success of our national literacy strategy, which includes a training program for all primary teachers in the best practices in teaching literacy to students aged five to 11 and a commensurate set of accountability measures. With this strategy, we have seen a very steady increase in test scores through last year. In 2001, around 150,000 more pupils achieved high standards at age 11 than in 1997. Through our national numeracy strategy, we brought about a similar rate of progress in mathematics, but with more undulations on the way. In 1998 we incorporated a new mental-arithmetic element into the test, which caused a dip in scores that year, but this was followed by a large rise in 1999. Since 2001, performance has plateaued at the new higher level, but we expect further improvement in the next two years.

The most impressive aspect of these two primary-level strategies is not so much the improvement in the average scores, but that there has been improvement throughout the spectrum. The biggest improvements in literacy have been at the high end of the spectrum, at level five, even though there was no target set for achievement at that level. In mathematics, similarly, although we set a level-four achievement goal, there has been very steady progress in reducing the number of students scoring in the below-basic category. Moreover, progress in both literacy and numeracy has been fastest in the most disadvantaged areas of the country. In other words, we are narrowing the achievement gap. Progress at the secondary level has also been impressive, though more incremental. But we expect to exceed our target for 2002.

So how are we doing? Test scores are rising very substantially and quite rapidly at the primary level and incrementally at the secondary level. They are rising throughout the ability range, throughout the age range, and the overall performance in the system is progressing. As a result of these reforms, the number of failing schools in the system has declined.

# WHAT HAS WORKED?

The first thing that worked—probably the most important thing—is that there is an underpinning framework for continuous improvement.

<sup>&</sup>lt;sup>1</sup> These figures include England and Scotland but not Wales, for reasons I have never understood. But I presume this does not affect England's position very significantly.



Figure 1 High Challenge–High Support Model

This framework took five or 10 years of experimentation and continuous reform to put in place, but it emerged in its current form in 1998–99. We call the model high challenge–high support. (See Figure 1.) We set high standards both for the national curriculum and for school inspections, against which all schools are measured. We then devolved as much money as possible to the individual schools. Ninety percent of all our school funding is in the hands of schools themselves, to be deployed by the head teacher or the principal. They hire and fire staff. They choose how many teachers, how many other staff, and how many computers they want. Every year we process data that enable schools to see how well they are doing compared to other schools in a benchmark group and compared to all other schools in the system. We also use the data to identify successful practices and then invest in professional development related to best practices, so that best practices get transferred around as rapidly as possible.

Having given schools clear standards, greatly increased funding, comparative data, and best practices, we hold them accountable by publishing their test-score performance annually and by inspecting them on a four- or five-year cycle. This accountability system leads to rewards for the successful, assistance for those who are working hard in difficult circumstances, and consequences for those who are evidently, according to the data and the inspection, underperforming. Underperformance may lead to a school being closed and the pupils transferred elsewhere, or it may lead to the school going through an improvement program.

This framework is undoubtedly very similar to many of the standards-based reforms in the United States. It provides an underpinning for a series of national strategies that have been developed, which reinforce it and accelerate its impact. In addition to the literacy and numeracy strategies at the primary level mentioned above, we now have a middleyears extension of those literacy and numeracy programs, which is at the end of its first year and we hope will deliver results this summer and next year. We have a program called Excellence in Cities designed to assist inner-city schools to collaborate in facing their problems. We have also been very tough throughout on school failure because we always put pupils first. We have a rapidly increasing number of "specialist" secondary schools that have centers of expertise in a particular subject but then share it with a network of schools. The evidence suggests that each of these programs is having a positive impact.

Investing in education is part of what works. By 2006, we will have had eight consecutive years of spending growth at roughly 5 percentage points above inflation. Not only is the total sum of money that we are spending increasing, but the amount that we are delegating to the schools year-on-year is also rising. Significantly bigger school budgets enable these improvements. But the key to Cathy Minehan's point about scale and speed is high-volume, high-quality professional development at the same time as clear accountability. If you can effect high-volume, highquality professional development, you can actually do large-scale reform pretty quickly.

### WHAT HAS NOT WORKED?

Quite a lot of things have not worked. There is not enough good behavior in secondary schools. A small but significant proportion of children aged 14, 15, and 16 do not attend school every day. Our vocational education compares very badly with other parts of Europe and parts of the rest of the world. We have a dropout rate at age 16 that is much greater than we would like when we look at international comparisons. We have introduced quite a lot of out-of-school education, but the connection between school and out-of-school is not quite right from the point of view of the student. Though our Centre for Gifted and Talented Youth has just been launched, we do not yet have real excellence in serving the most talented students, which seems to me a key element of twenty-first-century public education. And lower socioeconomic groups do not have sufficient access to universities.

Having devolved so much money to schools and given head teachers

so much responsibility, leadership development has become the fundamental building block of the next stage of reform, and we do not have that working as well as we would like. We have quite high volume, but it is not yet of sufficient quality to sustain the changes we want. Other issues follow from that: If we get the leadership right, we will generate the capacity to innovate and enable the system itself to lead the next wave of change.

To summarize, we are making progress, but we would like much faster progress at the secondary level than we have. We would like to see much greater belief among parents in the excellence of the education system. Too many of our reforms have been too clumsy and bureaucratic in their implementation, so that the reforms feel like impositions. We have not yet mastered this knowledge-transfer element. Too many teachers feel overworked and confused, although many of them also feel very proud of what they have achieved.

Despite the positives, this is quite a long list of negatives. Our reform agenda certainly is not finished yet; major challenges remain.

### WHAT'S NEXT?

In his book, *Good to Great: Why Some Companies Make the Leap . . . and Others Don't*, Jim Collins writes, "I am not suggesting that going from good to great is easy . . . . I am asserting that those who strive to turn good into great find the process no more painful or exhausting than those who settle for just letting things wallow along in mind-numbing mediocrity" (2001, p. 205). I think our challenge is the same. We have to go from good to great, and we do not think that it will be easy, but we prefer it to the alternative.

So far I have given you a semi-official report on what the British government thinks about its own education reform. I will now discuss the reform agenda for the next five to 10 years, not just in our country, but in the education systems of all developed and possibly developing countries as I personally see it. Take my remarks in that context. At the beginning of *Huckleberry Finn*, Mark Twain observes that *Tom Sawyer* is "mostly a true book; with some stretchers." From here on, my remarks have a few "stretchers."

There are two goals for what lies ahead. The first is the fullest possible development of talent, wherever it can be found. There is a war for talent, globally, nationally, company by company, not to mention soccer team by soccer team. From a soccer team's point of view, there are only three ways to get talent. One is to work with and coach the players you have. That is important and will keep you going for a while. Then you can buy it, but as you know, as in baseball and American football, true talent is very expensive to buy. Or you can grow your own. You can have a youth system that brings through, trains, and develops the sports people for your team of the future. The same is true for a company and true for a country. Each country needs a talent strategy that combines these three elements: growing your own (in other words, getting the education system to work well); buying talent (as, for example, California buys IT specialists from India); and of course, training and developing the workforce you have.

The second goal is equity. We have made real progress with the strategies that I have discussed. We have had the fastest improvement in the most disadvantaged areas, which is impressive and exciting. But if we are really serious, then, as we invest more in education, we should begin to identify learning difficulties early, we should spend money training specialists in identifying and dealing with these problems, and we should not give up until we have cracked the puzzle of how all these childrenall of them—can learn to achieve high standards. The talent and equity goals are not in conflict as is sometimes presented in the education debate. They go together. Equity provides a level playing field so that everybody has the basic building blocks. The provision of opportunity is a key part of equity: If you never find the chance to play golf, for example, nobody will ever know whether you could have been a great golfer. Once the talent is sought, discovered, and identified, we need a ladder of opportunity that allows development. The top end of the ability range in any field is as much an equity issue as those that are more commonly debated in education circles.

How will education systems accomplish these goals? I will list nine means of developing this agenda.

**1. Informed Professional Judgment.** The first is to reform the teaching profession. When I was a teacher in the 1970s, we turned up in a classroom, shut the door, and did our best to teach the children in the class. We had a textbook or a curriculum set by the school, and we made it up as we went along. Some of us did quite well some of the time, but only a few people, of their own volition, chose to go out and find out what was the best practice in pedagogy and what the research said they should be doing in their classrooms. Only once in my years as a teacher did the head teacher of my school see me teach. When he left, he said, "Thank you very much," and he never came back. He did not fire me though, and he would not have had the power to then. I was an uninformed professional. I could have become informed if I had sought out the information, but the system did not inform me. I used my professional judgment, but actually it was not really professional judgment—it was just amateur judgment.

In the 1980s, our governments became very frustrated as the international comparisons that were beginning to emerge showed the flaws in our education system, and officials began to prescribe some reforms. Some of them worked, and others did not. The reformers were uninformed, too. They decided to act, but nobody really knew what made for successful reform. The United States had *A Nation at Risk*, we had Margaret Thatcher, and the result was a growing interest in reform in both countries. At this point we had uninformed prescription. Out of that, some very important developments came, such as the National Curriculum, national testing, the inspection system, and the beginnings of the devolution of resources. By the early nineties something approaching systemic reform had emerged.

By the time the Blair government took office (and I joined the Department for Education and Skills), we knew a lot about how to prescribe changes. We had been learning about reform for 15 years. We wanted to prove that it was possible to implement system-wide change that delivered results rapidly because nobody believed it could be done. We set out to drive reform very quickly on a very large scale. We used informed prescription and made real progress, as the above results indicate.

But reform of that kind can take you only so far. It risks creating dependence and does not necessarily establish the foundations for the system to improve itself continuously. Keith Joseph, a British education minister in the 1980s, once said that the first words a baby learns are, "What's the government going to do about it?" This may describe the relationship now between government and teachers. As the problems develop, the teachers do not think, "How should we as an education system solve this problem?" They ask, "What's the government going to do about it?" Yet now that the government has devolved 90 percent of all the money to the schools, the money to solve the problems is in the schools, and so this is the wrong question.

The teaching profession and the government need to move toward a new phase: informed professional judgment. This means teachers who are driven by data and by what the data tell them. It means teachers who seek out best practices, who expect very high standards from all of their students, and who, when a student does not achieve high standards, asks the question, "How do we change what we do to enable that student to achieve high standards?" The shift over time is from a knowledge-poor to a knowledge-rich education system. We have a mass of very good data in the system. Shortly, we will have every individual pupil separately identified by a unique number in a national database, and we will be able to track different groups of pupils and individuals through the system. The combination of professional judgment with good data and a rich knowledge base will enable the era of informed professional judgment. This is a challenging but also an exciting concept for teachers, requiring a much higher level of discipline in relation to best practice than in any previous era. 2. Time: The Hidden Variable. Since 1988, we have reformed absolutely everything in the British education system except the one thing that dictates what people do all day: how they use their time. The way teachers use their time, the way students use their time, and the way the other adults in the school system use their time are the hidden variables in reform. Although we keep giving money to schools and allowing them the opportunity to change how they use their time, in fact, most of them, most of the time, just repeat the timetable they had last year, and the year before and the year before that. One issue for reform therefore is, how do you change the way people use their time? How do you, to put it in business terms, re-engineer the process, rather than use the extra money to do a bit more of the same process? And especially how do you persuade school leaders and teachers to undertake this re-engineering for themselves since prescribing it will not work?

Some leaders have begun to do some serious thinking about this. One head teacher in London said to me last year, "After four consecutive years of growth in the school budget, I suddenly realized doing more of the same wasn't the answer—we could transform our working patterns." So there are some schools at the cutting edge. The Milken Foundation in California has also published an excellent document showing how you could restructure a staff schedule with a 4 percent increase in an average high school budget. The schedule includes time for induction-year, associate teachers to do professional growth, planning, and mentoring, and includes blocks of time in a master teacher's timetable for mentoring, modeling, and coaching. All of this is perfectly possible in our system and would not require any extra money. But many schools are not actually rethinking time yet.

**3. Best Practice.** The third means is best practice. People who work in education have a tendency to shudder when you say "best practice" because they—we—were brought up in the period of uninformed professional judgment, when the core value of teachers was that it was up to each individual to make decisions on what should happen in his or her classroom. In fact, even after a period of prescription, the system has not yet become prescriptive enough about detail, which is what makes the difference in pedagogy, as it does in many other activities. When we trained teachers to teach literacy, they asked, "Can we be flexible about this?" We replied, "Well, this is the model that works. If you are flexible about it, it won't be pure, it won't be based on research, and then it won't work, and you'll tell us the program was a bad idea." Too much flexibility too soon has undermined many programs. Compare the impact in the United States of the flexible Coalition of Essential Schools with the much more powerful effect of "Success for All."

It is interesting to compare parallels in business, such as the 16 steps to checking into a hotel and then checking out again, published recently in the *Harvard Business Review*. For each of the 16 steps, there is a very

specific best practice (see Iacobucci and Nordhielm 2000). The article is interesting for its detail and for incorporating sources of best practice from outside the sector of interest. How often do we do that in education? For example, one of the things we know from education research is that a very good summary of a lesson by a teacher makes a huge difference in learning gains. But maybe the best person to coach teachers on how to give a good summary is not another teacher. It may, for example, be a very good chair of a business meeting. It may be somebody in a wholly different part of the economy. We are not yet obsessive enough about best practice or looking for it in all the right places—a thought that leads to my next point, about benchmarking.

4. Benchmarking. Benchmarking in education is becoming global. I do not know how much debate there has been in the United States about the PISA results published in December 2001. In our country there has been very little debate about it because it was good news, so needless to say, our press hardly reported it. But in Germany the results have caused a crisis. They have had the same effect that Sputnik had in America in the 1950s or that *A Nation at Risk* had in 1983. Two weeks ago, the front page of *Der Spiegel*, the best-selling news magazine in Germany, read, "Going Crazy, the New German Education Catastrophe." It was the second front-page story on education in *Der Spiegel* within a few months. *Der Spiegel* is currently running a whole series on education, which is a central issue in the German election campaign. This is a system in crisis. Those of us involved in PISA, which includes the whole OECD, will find increasingly over time that its results will set the agenda. They will set us on the search for best practice wherever we can find it, not just about pedagogy, but also about system design, about specific reforms, and about processes for implementing reform.

**5. Transparency.** In England, we publish performance tables showing the performance of every school in the country every year. Parents are very interested, and schools are very interested. But maybe this is only the beginning. People in the United States, particularly in the Federal Reserve Bank, do not need to be reminded of the need for transparency and trust in the few months after the Enron/Arthur Andersen disasters. Transparency and trust go together. People are going to keep investing in education at the levels they are now investing only if they see where their tax pounds, or their tax dollars, are going and what results they are getting. Where is the money going? What outcomes is it delivering? There will be pressure from taxpayers for ever more transparency in the system. There will also be pressure from consumers over time—the parents, or the students as they get older. Many educators see transparency as a threat because it means you cannot hide failure. Indeed, this is the main benefit—once a problem is out in the open someone has to be on the case. But transparency also challenges government. We publish so much data now in our country that every government policy is judged by its impact

on results. This is a very good thing. If a policy appears not to be working, then we have to defend it, argue it through, fix it, or stop it. So, transparency cuts both ways. Evidence of the British government's continuing drive for transparency in the public services is its response to a recent inquiry into unnecessary deaths at a Bristol hospital: "From 2005 results will be published annually for each centre and for each cardiac surgeon on a rolling 3 year basis.... This is just the first step to publishing more information on individual consultant outcomes" (Department of Health 2002, p. 117). The question arises then, when will schools begin to report to parents and students on the performance of individual teachers at the school? This is not government policy, but it is the trend in business and in healthcare, and I suggest it will become the trend in education too at some point.

6. Funding. Funding is a central issue for every public service. It is related to transparency because people want to know where their money goes and what they get for it. Frederick the Great of Prussia, one of Europe's most influential monarchs, said that finances are "the nervous system of the country: if you understand them you will be the master of everything else." Watching where the money goes is a key issue. And we need to become cleverer about the way we put money through education systems. The devolution of funding to schools in our country has been a huge step forward. But what about other funding, funding for out-ofschool programs, for example? Why not a voucher or tax credit for funding out-of-school learning activities of children from relatively poor backgrounds, as I proposed in my book The Learning Game (1997)? This way, we would provide the social-capital benefits of out-of-school learning for all children-benefits available right now only to children whose parents have the will and the means to provide them. Given the recent growth in funding of out-of-school learning and the expansion of provision, this is a question now of how the money flows, rather than finding large additional sums.

7. Elegance. Between the eighties and the nineties, we made a lot of progress in understanding how to achieve education reform, and we got better at it, but none of even the most passionate advocates, among whom I include myself, would call the reforms of the last decade elegant. But we will have to become elegant because we cannot keep creating so much "noise" with the way we do reforms. We have to get much cleverer in implementation. This requires government to learn rapidly and effectively from its own experience and that of others, and then to apply that learning. This is not a question of compromise; it is a question of clarity.

If we seek elegance, here are four questions we could ask about any reform before embarking on it:

• Does the particular reform fit well with the overall strategy both in concept and in timing?

- Second, are we getting the maximum change for the minimum investment? In other words are we maximizing leverage? Our national literacy strategy costs £80 million a year. But we spend £6 billion a year on primary teachers' salaries, so it is a good investment, a gearing ratio of about 1:80.
- Third, does the investment in the reform also strengthen the intellectual and social capital of the system as a whole, for example, by enhancing teachers' skills? If designed properly, every reform can achieve its objectives and simultaneously build capacity.
- Fourth, which levers should be used to implement the strategy? Do we really need to create new levers, or could we adapt the ones we have already? Are we using traditional bureaucratic or regulatory means of bringing about change?

I think the question of elegance is going to be a key part of reform because we are going to need faster reform at a larger scale, and if we do it with clumsy tools, we will drive reform into disrepute and teachers into despair.

**8.** Collaboration. One of the big paradoxes in education reform relates to the source of innovation. We have put so much money into the schools so that they can innovate and become the leaders of reform. The goal is that instead of the reform being driven from the center, schools take over the leadership. But it turns out that an individual school, even with control of 90 percent of the funding in the system, is not necessarily innovative. Most individual schools turn out to be rather conservative and risk-averse.

In the next phase, we will need to build schools into networks. Just as individual schools do not innovate, neither do hierarchies and bureaucracies. But networks may. Individual schools or organizations from outside the school system could lead networks, and government should enable them to do so. It might be, for example, that Chester Finn's virtual charter schools could become a source of innovation. Networks will form not on the orders of government, but because government has created the circumstances that will allow them to happen.

**9. Customers.** Finally, the reason education systems required powerful accountability systems in the nineties was that school systems until that time were not truly responsive to students and parents, though they sometimes spoke that language. They were not actually really meeting students' needs nor were they seeking out what the aspirations of students and parents really were. On the contrary, it was the interests of producers that came first. Accountability systems redressed that balance. The key to moving to lighter accountability systems with greater precision is building the customer into the equation. If schools were genuinely

accountable to parents and students, then accountability systems in their current form would turn out to be much less necessary than at present.

#### Conclusion

In conclusion, if we can get to a system where the data motivate schools to improve teaching and learning continuously while simultaneously seeking to understand profoundly the needs and aspirations of their students, then we will have a system that is one of informed professional judgment and is led by innovators in the teaching profession. Accomplishing this will not be easy. No one anywhere really knows how to do this yet. We are looking at a whole new frontier—as when Jefferson sent Lewis and Clark out to find a navigable route from the East Coast to the West. We have reached Kentucky, perhaps, but we do not yet know what is beyond the Mississippi. We have challenging questions ahead. Research and, eventually, policy will need to address these questions, because in the long run, the capacity to bring about rapid, continuous, large-scale education reform, and therefore raise standards of student performance to unprecedented levels, is fundamental to all our social and economic prospects.

#### References

- Barber, Michael. 1997. The Learning Game: Arguments for an Education Revolution. London: Indigo.
- Collins, Jim. 2001. Good to Great: Why Some Companies Make the Leap ... and Others Don't. New York: HarperBusiness.
- Department of Health. 2002. Learning from Bristol: the Department of Health's Response to the Report of the Public Inquiry into Children's Heart Surgery at the Bristol Royal Infirmary, 1984–1995. Command Paper CM 5363. London: The Stationery Office.
- Iacobucci, Dawn and Christie Nordhielm. 2000. "Creative Benchmarking." Harvard Business Review 78 (3): 24–5.

# IMPROVING EDUCATIONAL QUALITY: How Best to Evaluate Our Schools?

## Eric A. Hanushek and Margaret E. Raymond\*

It is difficult to be against accountability for public schools. Schools are creatures of state and local government, with all the associated expectations of performance and oversight. The importance of education for individuals and for society is unassailable. But many believe that U.S. schools are not contributing as much as they could and are not competitive in comparisons with those of other countries. Thus, the desire to hold schools responsible for outcomes is natural.

The disagreement comes, however, soon after people acknowledge the importance of school accountability. How should performance be assessed? Is providing information on student outcomes sufficient to get improvement? Should there be explicit sanctions and rewards for students and/or schools? Do unintended consequences overwhelm the intended consequences?

This paper considers some of the basic features of school accountability systems and assesses both the incentives for change that are imbedded in these systems and the existing evidence we have about behavior under different systems. The essential features that we consider are focus, scope of measurement, design, and incentives. "Focus" describes the mix of factors examined within accountability systems. "Scope" considers the extent to which the accountability system captures the full range of school activities for each of the factors under review. "Design" refers to the specific approaches to measuring the schools'

<sup>\*</sup>Paul and Jean Hanna Senior Fellow, Hoover Institution, Stanford University, and Senior Research Associate, Green Center for the Study of Science and Society, University of Texas at Dallas; and Research Fellow and Director of CREDO, Hoover Institution, Stanford University, respectively. This research has been supported by grants from the Packard Humanities Institute and the Smith Richardson Foundation.

contribution and the precision of these measurements. Incentives are created by the interplay of these three aspects of accountability systems and illuminate the ways schools will react to these initiatives. Some prior analyses provide reasonable tests of various incentives in action, and we provide some new evidence about the early impact of accountability systems.

The existing accountability discussion is surprisingly vague both in terms of what is being done and what should be done. Since much of the work to date has focused on single systems or isolated attributes or effects, it is hard to make informed judgments about accountability as a policy. A preliminary step of the analysis is to provide a description of where accountability stands in the United States today. This is essential for any evaluation of where accountability systems should be going.

### The Focus of an Accountability System

In almost all the states that have implemented school accountability to date, the overriding concern is the achievement of students. In contrast to policies of earlier periods, the chief focus of accountability is results, not effort. Most of the enabling legislation explicitly states that the purpose of adopting school accountability systems is to reflect student achievement outcomes and school performance.

That having been said, states have made differing choices in program design that have narrowed the range of outcomes and that frequently have involved other school characteristics. The premise of our discussion is that schools will respond most strongly to data elements that are included in the program, and that those aspects of schooling that are not included or measured will be de-emphasized, distorting school responses.<sup>1</sup> We consider a variety of aspects of this issue below; here we concentrate on the kinds of measures included in accountability systems.

In early 2002, CREDO surveyed each state in an effort to understand the details and effects of accountability systems, an area in which prior data have been very scarce.<sup>2</sup> Most states provide information on the districts in their state, and many have now taken this information down to the level of individual schools. Just providing unprocessed information is, however, considerably different from developing aggregate performance measures and putting rewards and sanctions into place. We

<sup>&</sup>lt;sup>1</sup> The incentive effects of choice of measures to be used in rewards have been extensively considered in the economics literature about optimal contracts; see, for example, Baker (1992, 2002) and Dixit (2002).

<sup>&</sup>lt;sup>2</sup> See CREDO (2002) for full details and citations of the analysis. CREDO, formerly the Center for Research on Education Outcomes, is an independent research unit at the Hoover Institution and has the mission of promoting and assisting in the evaluation of education programs and policies.

Table 1

Input	Process	Outcome	
Teacher Attendance Rate	Student Attendance Rate	State Achievement Tests (various grades)	
Condition of School Facilities and Grounds	Percent of Students Taking		
	State Test	College Entrance Exam Scores	
Number of Computers	Principal Mobility	Drop-Out Rate	
Course Offerings	Student Mobility	Graduation Rate	
Number of Non-Credentialed Teachers School Crime Rate	Teacher Mobility	Number of Students in	
	Year-Round School Status	Advanced Courses	
		Parent/Community Satisfaction	
		Percent of Students Passing End-of-Course Exams	
		Percent of Students Passing High School Exit Exam	
		Retention Rate	
		Suspension Rate	
Source: CREDO (2002).			

Variables Used or Proposed for Use in Accountability Systems, by Type

distinguish between simple "report cards" and accountability systems by the presence of aggregate measures that can be assessed against a standard, and by the use of rewards and sanctions related to measured performance.

While emphasizing student performance, another distinction between states is that in many cases the accountability legislation calls for the inclusion of other factors that do not measure outcomes. A significant number of states rely on a mix of process and input measures as well as outcome measures. With such a blend, those states hold schools accountable for the way students are taught in addition to considering the outcome of those efforts.

The incentives of hybrid systems are ambiguous: A school could be rewarded for improving its procedures, even if it does not result in additional student achievement. In contrast, an exclusive outcomeorientation creates incentives for schools or districts to direct resources appropriately in order to maximize the outcomes being studied. Outcome measures illustrate most clearly the degree to which schools are achieving the educational goals for their students.

Table 1 lists measures that have been incorporated into school accountability formulae or have been proposed for adoption in legislative bills, divided according to whether they are input, process, or outcome measures. Only 10 of the variables in Table 1 are outcome measures. Six

#### Table 2

Strength of Relationship to Student Achievement of Accountability Variables in Use or Proposed for Use

Weak	Moderate	Strong	
College Entrance Exam Scores	Condition of School Facilities	Drop-Out Rate	
Course Offerings	and Grounds	Graduation Rate	
Number of Computers	Percent of Students Taking State Test	Number of Students in Advanced Courses	
Number of Non-Credentialed Teachers	Student Attendance Rate	Percent of Students Passing	
Parent/Community Satisfaction	Teacher Attendance Rate	End-of-Course Exams	
Principal Mobility	Year-Round School Status	Percent of Students Passing High School Exit Exam	
School Crime Rate		Retention Rate	
Teacher Mobility		Student Mobility	
		Suspension Rate	
Source: CREDO (2002).			

are measures of school activities and are classified as process variables. Six are measures of inputs.

State systems that consist only of student test scores in each school have the virtue of being exclusively outcome-focused. This is not to say that they are perfect, because they can still be rather narrow in their coverage. Expansion of those systems with other outcome measures might add depth to the outcomes picture and still retain clear incentives for schools. However, where states include input or process measures, the strength of the association between the new measure and student achievement determines the degree to which the incentives are dulled. If the relationship between these other factors and student achievement is strong, then the combination would be less compromised than if the strength between them were weak. In short, the potential incentive rests on the degree of alignment between the measured factors and the outcome of interest.

Here we provide a summary of what the education literature indicates about the strength of the relationship of each variable to student achievement. Table 2 classifies each variable in one of three ways based on how strongly it aligns with student achievement and on the weight of existing empirical evidence. If the relationship has not been studied or the evidence is weak or inconclusive, we considered it to have "weak" support for inclusion in a school accountability system. If there is conclusive evidence about a variable but the estimated impact on student achievement is low, we concluded the strength is "moderate." If the

Achievement <sup>a</sup>					
Rating	Input	Process	Outcome		
Weak	4	2	2		
Moderate	2	3	0		
Strong	0	1	7		
	ef verdelele e freue Telele d				

Table 3 Pattern of Classification Variables by Strength of Association with Student Achievement<sup>a</sup>

<sup>a</sup> Value is the number of variables from Table 1.

Source: Authors' calculations based on Tables 1 and 2.

conclusive research showed a close and robust association, it was designated as having a "strong" relationship. (Note that we consider direct measures of student achievement tests as obviously a strong measure of outcomes and thus do not include them in this part of the analysis.)

The resulting classification, when put in terms of the underlying type of measure, shows an interesting pattern, as revealed in Table 3. The input variables were found largely to have a weak relationship with student achievement. Process variables have more mixed relationships to student achievement. Of the three types of variables, the outcome variables show the strongest association with student achievement, with two exceptions. The outcome measures of Parent/Community Satisfaction and College Entrance Exam Scores are weak indicators for the same reason other kinds of measures are weak—they show insufficient correlation with overall school quality. Public opinion research has documented a constant positive regard by parents for their children's schools despite actual differences in performance. College entrance exam scores, while providing some information, are self-selective and reflect only a segment of the student body of a school and thus can provide misleading summaries of the school outcomes because the sampling fractions are generally unspecified.

While we have no formal tests here, we assert that states whose accountability measures are more closely aligned with student outcomes deliver more consistent incentives to their schools. If a school faces consequences—good or bad—for teachers' professional development and for academic achievement, for example, the school will seek to allocate resources and place emphasis on both these dimensions of its operations. In its simplest form, we expect decisions to be made in accordance with the school's ability to change the measured factor, the cost of changing it, and with the reward (or punishment) associated with the change. This response is natural. However, if the two dimensions are not strongly related, a school could end up working at cross-purposes as they, say, pursue superior opportunities for teacher training and also work to improve student learning. Taking teachers out of the classroom for their professional development activities may actually work against improved learning in their classes.

Since many of the inputs and processes are more concrete than the outcomes—we know how to order more computers or to deliver new programs—they are the low-hanging fruit on the accountability tree. Any elements that are associated closely with the more difficult and desirable objectives of student achievement reinforce the incentives that prompt schools to take corrective action. However, since the majority of the input and process measures currently in use do not meet that standard, they dilute the strength of the output incentives and generally weaken the system.

#### SCOPE OF MEASUREMENT

The scope of an accountability system highlights the breadth of focus that a state elects to adopt. The scope of the accountability system will have an effect on how strong the incentives are and how much latitude or slack schools retain to minimize the impetus to change. Interviews conducted with state officials in 2001–02 suggest that the strength of the incentives is directly related to the comprehensiveness of the program that a state implements (CREDO 2002).<sup>3</sup>

States appear to have been influenced in their choice of scope by several factors. There may be resource constraints that necessitate a narrow focus. Political dissent about implementing accountability in any form may require concessions in the breadth of the program. Some states may wish to proceed cautiously in order to be able to adjust incrementally as the program matures. And there is some evidence that states may even have genuinely different theories about their appropriate role in gauging school performance. CREDO (2002) mentions all these factors as explanations of the varying structures implemented by state officials and of the differences that are observed across systems.

The implications of many of these larger issues are difficult to assess, in part because they have not been clearly linked to measurable outcomes. We can nonetheless look at some of the factors that enter directly into school testing programs.

The clearest indication of differences in scope can be seen by

<sup>&</sup>lt;sup>3</sup> Much attention has been given to the potential implications of narrowly focused accountability systems. Most frequently this is raised with respect to the types of achievement-measurement instruments that are employed. For example, do they emphasize just "lower-order" skills or do they concentrate on items most easily included in standardized testing? But the debate also includes issues of whether concentration on basic cognitive skills drives out other elements such as citizenship development, character education, and the like.

Table 4

	by the Nulliber of Clade Levels Assesse	30 11 2001			
Minimum (Fewer than 5 Grade Levels)	Better (5 to 8 Grade Levels)	Best (9 or More Grade Levels)			
Connecticut	Alaska South Carolina	Alabama			
Georgia	Arkansas Texas	Arizona			
Hawaii	Colorado Utah	California			
Indiana	Delaware Vermont	Idaho			
lowa	Florida Virginia	Mississippi			
Maine	Illinois Washington	South Dakota			
Minnesota	Kansas	Tennessee			
Montana	Kentucky	West Virginia			
Nebraska	Louisiana				
Nevada	Maryland				
New Hampshire	Massachusetts				
New Jersey	Michigan				
New York	Missouri				
North Dakota	New Mexico				
Ohio	North Carolina	North Carolina			
Oregon	Oklahoma				
Wisconsin	Pennsylvania	Pennsylvania			
Wyoming	Rhode Island				
Source: CREDO (2002).					

Classification of States	s by the Number	of Grade Levels	Assessed in 2001

considering the number of grades of schooling that the accountability system covers and whether the included grades are sequential or discontinuous. Both aspects have an effect on the incentives that an accountability system produces and their effect on schools.

Table 4 presents the 50 state systems classified by the number of grades included in their testing system. Eighteen states include fewer than five grade levels, 24 states cover between five and eight grades, and eight states have nine or more grades. Even among states with the same accountability model, differences in the strength of the incentives will arise from differences in the scope of their performance focus.

Note that we are not able to judge the quality or breadth of the separate state examinations. We do record whether the tests are criterion-referenced (developed for the specific objectives of each state's schools) or norm-referenced (more generic tests applied across the nation). This division provides some information about the relationship between each state's testing program and its educational goals and standards. None-theless, it is a rather coarse cut across the testing programs.

Among states with more grade levels included in the school scores, differences remain. As reflected in Table 5, which shows the grades and types of assessments currently in use by states, the majority of states capture achievement from an erratic pattern of grade-level testing.

	Norm- referenced	Criterion- referenced		Norm- referenced	Criterion- referenced
Alabama	3-11	5-7	Montana	4,8,11	
Alaska	4,5,7,9	3,6,8,10	Nebraska	4,8,11	
Arizona	2-11	3,5,8,10,11	Nevada	8,10	
Arkansas	5,7,10	4,6,8,11	New Hampshire		3,6,10
California	2-11	2-11	New Jersey	4,5,8,11	
Colorado		3,4,5,7,8	New Mexico	3-9	
Connecticut		4,6,8,10	New York		4,8,12
Delaware	3,5,8,10	4,6,8,11	North Carolina	3-8,10	
Florida	3-10	3-10	North Dakota	4,6,8,10	
Georgia	4,8	4,6,8,11	Ohio		4,6,9,12
Hawaii	3,6,8,10		Oklahoma	4,5	5,8,9-12
Idaho	3-8	4,8,9-11	Oregon		3,5,8,10
Illinois	3,5,8-12		Pennsylvania		5,6,8,9,11
Indiana	3,6,8,10		Rhode Island	4,8,10	3,7,10,11
lowa	4,8,11		South Carolina		3-8
Kansas		4-8,10-11	South Dakota	2-11	
Kentucky	3,6,9	4,7,8,12	Tennessee	3-8	9-12
Louisiana	3,5,6,7	4,8	Texas		3-8
Maine	4,8,11		Utah	3,4,5,8,10,11	
Maryland		3,5,8,9,11	Vermont		2,4,6,8,10,11
Massachusetts		4-8,10-11	Virginia	4,6,9	3,5,8
Michigan		4,5,7,8,11	Washington	3,6	4,7,10
Minnesota	3,5,8,10		West Virginia	3-11	4,7,10
Mississippi	5-8	2-12	Wisconsin		4,8,10
Missouri		3-5,7-11	Wyoming	4,8,11	4,8,11

т	a	h	ما	5
1	α	D	Iе	0

Type of Assessments Being Used in States and the Grade Levels Being Assessed

Compare, for example, the states of South Carolina, Texas, Utah, and Vermont. All rely on test scores from six grades, but in South Carolina and Texas the grades are consecutive. This is not the case for Utah and Vermont, which both sample from three elementary grades, one middle school grade, and two high school grades. Clearly, states with consecutive grades have an easier time attributing changes in school scores more accurately to their own activities. Beyond that benefit, schools with consecutive grades face steady incentives across those grade levels, which we would expect to result in consistent attention to each grade on the part of schools. With discontinuous grades, the opportunity exists to focus more strongly on the grades under review.

While the pattern of test taking is likely to change with recent federal legislation on testing and accountability, the message at this point is clear. Most states do not have a broad and uniform assessment policy across grades. This disjoint nature of testing both affects how well information

can be used to judge school performance and alters some of the incentives faced by schools. It is to these latter points the analysis proceeds.

# **DESIGN AND INCENTIVES**

Concentrating on student performance as the key focus of accountability will obviously transform the practice of the past, when a majority of states provided just rudimentary information about their schools, often confined to a few measures of school resources and avoiding any indication of student performance (Hanushek and Raymond 2001). Even where states have created a hybrid system that combines input and outcome regulatory elements, student outcomes have become a major focus. Yet, the appropriate use of student outcome information is far from obvious. The ways that states compile student achievement measures into school scores and how they treat those results create very different pictures of school performance.

Most of the accountability systems have implicit or explicit goals underlying them. In many cases, the goals are multiple; for example, to improve student achievement and to narrow the historical gap in performance across racial and ethnic groups. Thus the design of a system serves as the vehicle for translating desired goals into incentives that motivate schools toward these goals and capture the results for review. To the extent that a design ignores one or more goals or creates conflicting motivations, the system that relies on that design will likewise distort incentives.

## **SUMMARY MEASURES**

The key to understanding the informational content provided by state systems is to examine the determinants of student performance and how those determinants are displayed within the accountability system in each state. As a foundation, prior work on the determinants of student achievement identifies student outcomes as coming from a variety of influences: families, friends, teachers, and schools. Moreover, a student's knowledge evolves and builds on past learning and on the individual's skills and abilities. How these various influences are recognized and accounted for dramatically influences the ability of state officials to discern the performance of schools and to provide clear incentives.

Accountability systems begin by testing a group of students in each school and then presenting information about school achievement. The actual measure of school achievement varies. The simplest measure is the average of test scores of the students in a grade or an entire school, although few states end up developing their accountability systems on just school-average achievement. Important variants include distributional information such as the percentage of students scoring above some specific level ("passing" or "proficient"). These variants introduce important elements into accountability systems, but for now, we consider just the average performance measures. Virtually all states, whether they provide just report card information or instead develop accountability structures, report average achievement as one of the components of the information given.

#### Status Model

The status model simply uses the average performance of students as a measure of the outcomes in each school. (While it is more important later, we do not distinguish at this point between systems built on calculating grade averages as opposed to school averages).<sup>4</sup> The first point from this is obvious: If the main purpose of the accountability system is assessing the performance of the school, the average test score does it very imperfectly. In addition to school performance, the average achievement will incorporate all of the current and historical inputs to achievement including not only school but also family background and random errors. With the status model, it is not possible to factor out year-to-year changes in student-body composition, or grade-to-grade changes in instructional design or teacher quality. Thus, the simple average score indicates the level of student performance but cannot pinpoint the source of that performance. Despite these imprecise measures, schools are treated according to the result, for better or worse.

This basic confusion between average student achievement and the contribution of schools is well known, and some state accountability and reporting systems provide additional information that might be useful for adjusting these scores to get closer to the impact of schools. For example, some states either provide data on family backgrounds (such as rates of free lunch participation or racial compositions of schools) or describe achievement for reference groups of students judged to have similar family backgrounds. While these measures are usually available, they generally act merely as an external reference, but do not influence the results of the accountability calculations. Thus, these approaches highlight issues of accurate estimation of school performance, because they likely do not adequately identify family differences or cohort differences and they do not capture prior factors that affect current achievement. Nor do they allow for any measurement errors in performance. Most of the attention has focused on ways of trying to allow for differences in the

<sup>&</sup>lt;sup>4</sup> For average performance the distinction is unimportant, but a variety of state reward systems are based on such measures as the percentage of students passing a grade-level test. In those, performance requirements or rewards based on separate grades imply different incentives and constraints compared to school-based systems.
nonschool factors, but existing efforts have simply produced imprecise results, leaving considerable uncertainty about interpretation of scores and little way to separate out the value-added of the school.

One other aspect of status models is important—the relationship between goals and incentives. An underlying explicit or implicit element in most accountability discussions is that schools have systematically left minorities and disadvantaged students behind. In reaction, explicit goals of narrowing and eliminating the existing gaps have been translated into status accountability models built on unadjusted aggregate scores. This confuses goals with the incentives of accountability systems, because each school finds that incentives include aspects of performance that it does not control. Put another way, if one school has students who come to school with poorer preparation than another, that school must meet a higher standard in terms of its value-added to student learning.

A variant of the status model considers performance just for separate grades, instead of aggregate school performance. While the approach is still cross-sectional in nature, and, therefore, vulnerable to shifts in student composition, it provides a more precise focus on school inputs. The approach can help to provide schools with the ability to distinguish between school inputs and student variation. The effect from student migration will still exist, but cohort effects will be seen as they move across grades. With stable programs and teachers, teacher effects will persist over time.<sup>5</sup>

The grade-level variation of the status model of accountability also is used when testing does not cover the range of grades. If, for example, testing is done only at the fourth grade, the accountability system would feature just that grade.

### Status-Change Model

The status-change model tracks the average student achievement of a school over time. The idea is easiest seen in terms of an example. The status-change score for a school that has a common examination at a specific grade, say third-grade reading, would appear as the change in the average third-grade reading result between the 2000 and 2001 school years. The status-change model is often calculated for an entire school by aggregating the performance across tested grades.

The status-change model is by far the most common approach to assessing what is happening in schools. Change scores frequently factor

<sup>&</sup>lt;sup>5</sup> Note that the interpretation of year-to-year grade or status changes depends crucially on which information is used. If looking at just the difference in performance across cohorts of students, the relevant school effect is the change in school quality. If levels of performance are calculated at each year, information about the level of school quality inputs can be obtained.

heavily in reward systems, but they are treated in a wide variety of ways: Examples include absolute levels of change, percentage increments of change, and change relative to an external standard.

The most common interpretation, regardless of form, is that the status-change model provides a measure of the change in performance of the particular grade or school. Thus, for example, states may have goals or rewards related to the "progress" that is measured by the status change. Indeed, recent federal legislation also incorporates change in testing and accountability requirements. Does the accountability system built on status change provide biased estimates of performance improvement that systematically diverge in one direction or another? Are the errors so large that they mute any incentives for schools to do better?

Even if the student body of a school is identical across years, the status-change model is still comparing two different groups of students. Thus, status change has three primary components: the difference in school quality across the two years; the difference in family background and other nonschool factors between the two groups of students; and the average difference in any idiosyncratic errors affecting achievement. Just like the status model that relies on the level of average achievement, the status-change model completely entangles school performance with student-background differences and measurement errors. The best interpretation would be that, if variations in quality improvements across schools were large relative to differences in the other factors, changes in grade or school performance would dominate the changes. But, there is little existing evidence that would support that interpretation.

The situation is, however, even worse than many believe because of the dynamics of student populations. The mobility of the U.S. population has important implications for schools—not only for the way they teach students but also for their accountability systems. The U.S. population moves surprisingly frequently. From a recent Current Population Survey, we find that only 55 percent of students live in the same house over a three-year period, and this falls to about half for disadvantaged students. Moreover, residential mobility is often related to significant changes in family circumstances such as divorce or job loss and change. In growing states the mobility rates are noticeably higher. The average annual student mobility across schools in Texas, for example, exceeds 20 percent (Hanushek, Kain, and Rivkin 2001) and in California the figure is 15 percent (CREDO 2002).

The implications of mobility for the accountability approaches are clear. As mobility increases, differences in the backgrounds, preparation, and abilities of two groups of students compared over time will influence differences in aggregate performance in the status-change model. At that point, not only do current differences in nonschool factors enter the picture but historical differences also do—and mobility implies that two adjacent cohorts will also diverge in terms of the past schools they attended.

### **COHORT- AND INDIVIDUAL-GAIN MEASURES**

By shifting attention to the progress of students rather than schools over time, it is possible to gain substantial accuracy in the focus of the accountability system. Consider following the same students in a school, year to year, and calculating the improvement or decline for the cohort. The result is a new measure of school performance that has some superior characteristics. With a stable student body (that is, with no in- or out-migration for the school), the historical school and nonschool factors would cancel out (because they influence a cohort's performance both in the current grade and in the prior grade). The cohort-gain score would then reflect what the school contributed to learning plus any differences in idiosyncratic test factors across the two grades. The influence of family differences on current achievement growth rates would also remain, so that if, for example, disadvantaged students would be expected to have lower rates of improvement in performance than the more advantaged, such differences would remain confounded with school factors. Nonetheless, the cohort model would generally yield a closer measure of the school's contribution than the status model. The family background and ability factors that affect the cohort-gain calculations are ones that affect the rate of growth of learning, not the level. Thus, they would be expected to be relatively small.6

The final design that has begun to be used by states further refines the progress model by calculating gain scores for individual students and then creating school summaries by aggregating them by grade and by school. This approach provides the highest level of precision because it controls for family differences and differences in student body composition, and it isolates the year-over-year contribution of schools to student performance. Because it follows individual students, including in-migrants, it minimizes the effects of student variation. Cohort effects are still uncontrolled to the extent that a specific group of students may be brighter or duller than average (perhaps by design through exclusions). Since additionally it focuses on progress, the model can isolate the contribution of individual teachers, although no state makes such information public.<sup>7</sup>

The array of states under the different types of systems is presented

<sup>&</sup>lt;sup>6</sup> Some practicalities of calculations still remain. The primary question is how to deal with any mobility that might enter into the calculations.

<sup>&</sup>lt;sup>7</sup> Tennessee produces measures of individual student value-added, but they are not publicly released (Sanders and Horn 1994).

Systems in 2	2001	e Type OF A		USEU IN SCHOO	Induliy	
	Cross-Sectional	Student-Change Approaches				
Schoc Status-C	l Status or hange Model	Grade-L	evel Change	Cohort-Gain	Individual-Gain	
Alabama Arkansas California Connecticut Georgia Maryland Michigan Mississippi Nevada	New Hampshire New York Ohio Oregon South Carolina Texas Virginia West Virginia	Alaska Colorado Delaware Florida Kentucky	Louisiana Oklahoma Rhode Island Vermont Wisconsin	New Mexico North Carolina	Tennessee Massachusetts	
Source: CREDO	D (2002).					

Table 6 Classification of States by the Type of Analysis Model Lised in School Bating

in Table 6. The vast majority of states rely on cross-sectional measures and comparisons-even though these approaches generally have the least appealing properties. Only four states (Massachusetts, New Mexico, North Carolina, and Tennessee) currently emphasize student gains. The implications for incentives and results are developed in the next section.

# **INCENTIVES AND EVIDENCE**

It is useful to translate the discussion on the different accountability systems into hypotheses about the incentives introduced by each. We then provide a review of existing evidence about these hypothesized effects. It is important to bear in mind, however, that the recent birth of many accountability systems means that the existing evidence is thin in many crucial places. Indeed, the thinness of the evidence is one of the main points of this analysis.

Accountability systems are designed to increase the exposure of schools by revealing the quality of student performance. Two separate mechanisms operate: the public sharing of performance data and any directly legislated rewards and consequences.

Any school will prefer higher scores to lower ones, even if no explicit consequences follow the awarding of scores. Currently, apparently in the absence of much clear evidence, most parents think that their school is doing a good job (Rose and Gallup 2001). The sharing of accountability evidence has the potential for changing this, perhaps sufficiently to overcome the inertial positive regard for local schools. In the absence of direct consequences, one might expect any purely informational incentive to be small relative to organizational pressures to maintain the status quo.

Nonetheless, some general evidence on reactions of citizens (in the form of housing prices) to perceived school quality information exists (Black 1999; Weimer and Wolkoff 2001). Moreover, as discussed below, early evidence suggests that public disclosure of scores may in fact produce some strong incentives, both in terms of housing prices (Figlio and Lucas 2000) and other observable outcomes.

The second source of incentives from exposure of performance arises from any consequences that might be directly associated with the school scores. The rewards and sanctions that many states have built into their accountability systems create the motivation for schools to change behavior. At the same time, one does not expect these incentives to affect all schools equally. For example, schools that have many students scoring close to a threshold might be expected to alter their behavior more than schools with students further away from the established critical thresholds.<sup>8</sup> The interrelationship between the choice of a school-score model, the choice of thresholds, and the location of a given school relative to those thresholds is currently relatively unexplored, but it would be reasonable to speculate that no single design can provide equivalent incentives for all schools. Moreover, it is well known that incentives that emphasize crossing a specific threshold will generally lead to ever greater distribution in behavior.

The following sections consider in more detail the incentives under different accountability models. Within each section, we also provide a review of the existing evidence about the impact of the various incentives.

### **Cross-Sectional Approaches**

As delineated in the preceding discussion, the status model combines one-time scores of student performance into a single school score. Any change in scores from year to year generally is assumed to be a function of school influences. But, since the design does not recognize changes in the underlying student population, the model creates the incentive to include more positive student test scores into the school scores, that is, to adjust the relevant test-taking population.

A school can respond to disappointing assessments in two ways. First, it can adjust teachers, curriculum, and programs in an attempt to improve the teaching that occurs. This is, however, a difficult long-run proposition, made even more difficult in schools with high rates of staff turnover. A second, shorter-run strategy may result: to become more selective about the student scores that are incorporated into the school

<sup>&</sup>lt;sup>8</sup> See, for example, the parallel with past incentives employed in the experiments with performance contracting, where contractors reacted very openly to the notches in the contracts (Gramlich and Koshel 1975).

scores. The second approach could supplement or possibly replace the first. By weeding out students who are poor performers, the school score can appear to be improving even if nothing different is being done.

The dynamics of these alternative approaches are important. Take the example of a third-grade student from a disadvantaged background who arrived at school less well-prepared than the others in the school and who progressed at a slower rate each year through the third (that is, falls further behind over time). The status model compares performance of individual classes each year to the prior year's class. Thus, if testing begins in the third grade and the system has been going for some time, the school might exclude this slow student through placement in special education or by counseling the student to be absent on the day of testing. If the student is excluded in the third grade, the average of all remaining students would be higher than otherwise, and the school would tend to look better in comparison to the third grade in the prior year. But, the next year's comparison of third grades will be worse because the base comparison has been artificially elevated. Moreover, once the school has excluded a student, there is a continuing incentive to keep the student out of the testing. This continuing incentive puts some restraint into the system, because the school probably cannot increase the exclusion rate year after year. Moreover, since the potential importance of exclusion rates is widely recognized, the school is always at risk that regulatory changes may make it necessary in the future to bring some previously excluded students back into the accountability system.

While the largest effects of exclusion on the school ratings come in the first year of exclusion (when the cumulative effects to the current grade of low preparation plus slow learning are removed), there are some continued accountability benefits to the school from exclusion if the omitted students learn at a slower pace. The status model aggregates across grades, so the slower learning pace will be removed from the calculation of the school average for the student's fourth grade and beyond. The key element of this part of the dynamics is how much the rate of learning might be below average, as opposed to the absolute level of deficit that comes into play in the first year of exclusion.

While there has been widespread attention to such things as test preparation and cheating, these seem to be the clearest cases of one-time effects that are not sustainable after the initial introduction. Specifically, these practices may shift the level of performance in a given year, but, unless their prevalence increases over time, they will not show up in the school gains after the first year. Take, for example, efforts to teach all students how to fill in mechanical scoring sheets for standardized exams. Once students know how to do this—something that might inflate their scores through eliminating errors arising just from coding mistakes—it would not be expected to have any continuing effects on their scores as they progress through the grades. Similarly, any cheating on a given test must be repeated in subsequent years just to stay at the same level, but scores will improve only if the level of cheating is increased over time.

The choice of approach may be assumed to follow rational choice: School officials would select the action that they perceive to have the highest yield, given their planning horizon, budget, and appetite for risk. The preceding discussion highlights the fact that the largest gains from exclusions operate in the first year and that these decline or possibly reverse in subsequent years. Administrators may be very myopic or may have very short time horizons for their decisions, leading them to overuse exclusions in the first years of an accountability system. Regulatory restrictions frequently are designed in an effort to limit the ability of administrators to increase the use of student exclusions.

The grade-level change variation of the status model of accountability introduces some additional incentives. Some of the dynamics of exclusions are altered. But also there may be incentives to concentrate attention on the tested grade(s), say by placing the best teachers in the relevant testing grades.

Study of the exclusion rates of schools is one way to detect if schools are culling their student ranks prior to testing. Alternatively, one could examine the prevalence of parental waivers, with attention to which students are being held out. Finally, consideration of the effects of state policies on when students who change schools must be included in the new school's score could provide another perspective on exclusions.

Several studies have investigated whether schools appear to react to accountability through exclusions. Jacob (2002) considers the introduction of test-based accountability for Chicago public schools. He finds that the large increases in test scores after accountability went into effect were also accompanied by increases in special education placement and by increased grade retentions. Deere and Strayer (2001a, 2001b) and Cullen and Reback (2002) also find apparent increases in special-education placement with the introduction of accountability in Texas. Prior work on Kentucky by Koretz and Barron (1998) suggests no strategic use of grade retentions. Haney (2000) suggests that both grade retention and increased dropouts were key to improvements in Texas tests, although this finding is seriously questioned by reanalysis of the data. Both Carnoy, Loeb, and Smith (2001) and Toenjes and Dworkin (2002) find little evidence that testing led to the changes suggested by Haney. Carnoy, Loeb, and Smith also find that at least in larger urban areas lower dropout rates are associated with higher student achievement. The grade retentions are, however, short-run effects that do not provide lasting value except if the placement is educationally valuable. Figlio and Getzler (2002) concentrate on special-education placement after the introduction of a state accountability system in Florida. The most persuasive evidence is that placement rates increase relatively over time in grades that enter into the accountability system as opposed to those grades that do not.

Jacob finds that scores also appear to go up more in subjects that enter into the accountability system than in those that do not. This evidence is consistent with analysis in Texas by Deere and Strayer (2001b). The interpretation is not, however, entirely clear. Schools obviously appear to be responding to the accountability system—which is exactly what the system is supposed to accomplish. On the other hand, one might question whether the weights on different potential outcomes are appropriate. (Zero weight or not paying attention to specific subjects, for example, appears to provide very strong incentives to change the pattern of instruction.)

In each case, the analysis considers changes that occur around the time of introduction of an accountability system. In fact, the key element of most of this research is using the change in accountability to identify the effects on special-education placement rates and the like by finding breaks in the patterns of prior placement. Two things are important. First, there is very little relevant data for these analyses—breaks in trends or perhaps comparisons to trends of other schools (such as schools outside of Chicago and its accountability system) convey limited information. The validity of the interpretation depends crucially on whether or not other things are changing over time that could also affect the patterns of observed changes. Second, each of these analyses provides information just on the short-run immediate effects. Since the incentives change over time, it is important to understand what happens as these systems continue. Because of the recentness of introduction of accountability systems, little is known about the long-run dynamics.

Hanushek and Rivkin (2002) investigate the impacts of public disclosure of achievement performance. Specifically, before the Texas accountability system included direct consequences or sanctions for performance, the state made information on disaggregated student performance from the Texas Assessment of Academic Skills (TAAS) available to the public. They find that in the largest metropolitan areas, competition works to push up average scores.

Greene (2001a, 2001b) analyzes the Florida A+ program that provides exit vouchers to students in failing schools and finds that schools at risk of becoming sites of vouchers make unusually large gains. Carnoy (2001) reviews this evidence and suggests that the reaction to vouchers that Greene identified was more likely a reaction to information. Carnoy finds that similar studies for North Carolina and Texas (Ladd and Glennie 2001 and Brownson 2001, respectively) investigating what happens to failing schools show similar results—dramatic improvements in the year after identification. This occurs even though those states had no voucher threat.

On the other hand, Kane and Staiger (2001) suggest that a portion of the school improvement in North Carolina's failing schools may simply result from measurement errors in the examination scores. They demonstrate that small schools—where the error variance in aggregate tests will be larger—are much more likely to be found at the extremes of the school score distributions. If the measurement errors are independent over time, schools that realized a large error in one period would expect to receive a smaller one the next period, leading to a reordering of schools in the second year. Kane and Staiger do not, however, differentiate among the sources of error of the status model—family differences, teacher and school differences, and measurement errors.

The implications of grade-level versions of accountability have been less studied. Some of the prior work employs differences by grade level primarily as a method of identifying the behavioral effects of the system as opposed to being a focal point of the analysis. Boyd et al. (2002) do consider whether teacher placement responds to the specific grades that "count." They find that exiting from teaching does not appear related to testing regimes. While they have only indirect measures of teacher quality for the New York state sample (experience and quality of college), they do find some attempt in urban schools to place the more experienced teachers in the grades tested when new teachers entered a school.<sup>9</sup>

### **Student-Gain Approaches**

The two variants—cohort gain and individual student gain—produce an average score of student-performance change for a group of students. The distinction between the two in their pure form relates to the group of students included.<sup>10</sup> Student-gain measures allow for the school to isolate school inputs in much the same manner as the grade-level change model above. The superiority of the student-gain model over the grade-level change model lies in its control of student characteristics and in its focus on the level of school performance. Just two states as of fall 2001 (New Mexico and North Carolina) have employed a pure form that examines the same cohort of students year over year as they move

<sup>&</sup>lt;sup>9</sup> This evidence is not entirely conclusive about strategic behavior, however. If the grade-level accountability relies just on the levels of achievement in a grade (as most do), schools have an effect that accumulates over time. Thus, getting the effect of a good teacher is possible by placing that teacher in the grade being tested or in a prior grade where students would be better prepared for the material in the tested grade.

<sup>&</sup>lt;sup>10</sup> In pure form, the largest difference is whether the individual school gets information on the distribution of performance from the individual-gain calculations. An impure form, however, introduces some error in the cohort-gain measure. Specifically, a cohort gain can be calculated by taking the scores in a grade of all students in a school for two years and subtracting the average prior year scores for the previous grade. In this approach, people who exit between grades are included in the base but not in the current-year score.

through a school.<sup>11</sup> The focus on change instead of static performance lends itself to closer association with a school's efforts to improve.

The primary incentives inherent in this approach fall more on improving student scores by improving teaching and programs than for the status model. Exclusions could have an effect on measured performance to the extent that the exclusions eliminate individuals who would have a lower rate of learning. As noted above, however, this impact on the accountability score generally will be considerably less than the impact of exclusions on the status model, because it is only achievement growth and not achievement level that is important.

Since the group of students being examined is constant over time, the model ignores student in-migration. This outcome may interact with district decisions to set school attendance zones and the like—which would eliminate some students from the calculations. To date, no evaluations of the effects of cohort-gain systems on performance are available.

The student-level gain score model follows the progress of individual students and then creates a summary from the net change scores. Of all the models, this approach provides the clearest and strongest incentives for schools to concentrate on the school factors under their control since it minimizes student variation. It enables the fastest and cleanest feedback on any efforts the school undertakes.

With this model, the strength of the incentive will be a function of changes in student-body composition, but the effect will be smaller than for the cohort-change model. Even though student moves are known to affect scores negatively, as implemented, the school will have students for more than a year before their gain scores are included in the school score.

The model would create the inclination to exclude students who are poor performers. The school will know student-specific performance in the first year of examination and then can follow their progress through the second year, presumably providing information by which to prejudge which students would likely produce negative change scores. By avoiding a second-year test, the gain scores for those students could not be calculated or folded into the school score.

Richards and Sheu (1992) provide an early investigation of the South Carolina incentive system. This system, introduced in 1984, was a sophisticated accountability attempt that considered individual student-

<sup>&</sup>lt;sup>11</sup> Two aspects of the design of cohort-change systems are important. First, decisions must be made about exclusions of students because of mobility. Based on individual data, it is possible to use initial and subsequent scores for just individuals who start and finish the grade. In general, new entrants during the grade would be excluded from the calculations, but the data would not introduce errors from different groups of students. Second, across each year a decision can be made about whether to update the cohort to the group beginning each grade or whether to maintain the cohort originally identified.

gain scores and adjusted rewards for the socioeconomic status of the student body. They find that the reward system yielded gains, although modest, in performance of students (but did not affect teacher attendance, the other attribute of incentive focus). Interestingly, South Carolina subsequently moved away from this incentive system. Ladd (1999) investigates the sophisticated gain-score incentives in Dallas, Texas, during the mid-1990s. She finds that performance in Dallas improved relative to other large Texas districts, although the gains come from white and Hispanic students but not black students. Improvements in terms of student dropout rates and principal turnover rates also appear.

Deere and Strayer (2001a, 2001b) evaluate the impact of Texas incentives on a range of behaviors. They find evidence that schools tend to concentrate on students who are near the passing grade on the TAAS. Moreover, there is some tendency to concentrate on subjects that enter into the accountability system. The evidence also suggests some differential exclusion from testing. They specifically find some sharp increases in overall exemption rates for special education around the time when these exemptions became most important for accountability. (Note, however, that while the evaluation considers student gains, the Texas incentive system concentrates on overall pass rates.)

In terms of incentives, the objective of rewarding and punishing schools for their contributions to student learning are met in varying degrees by the alternatives. By far the most common alternative—the status model and its grade-level offshoot—provide information that is far distant from the value-added of each school. One aspect of this is the introduction of incentives to change school scores in ways that are unrelated to their learning outcomes. For example, increasing specialeducation placements or working selectively to decrease test taking can improve scores for a school by changing the rating group. Of course, some alterations work best in the short run—that is, in the year of their introduction—and would be much less effective in later years. The use of these approaches depends on the simple decision-making of administrators and is related to the costs, risks, and time horizons of the administrators.

# CUMULATED EVIDENCE ON INCENTIVES

Most accountability systems have been introduced very recently, so the history does not give much scope for analysis. Nonetheless, a variety of investigations have been undertaken recently and provide some, albeit limited, evidence. Table 7 groups these analyses by their focus and by the type of accountability system studied. It seems clear that schools do in fact respond to accountability systems.

Much of the evidence relates to "gaming" the system—actions taken in response to incentives but not directly related to improving perfor-

•	5
Cross-Sectional Accountability Systems	
Outcome Effects	
Direct Response to	Greene (2001a, 2001b); Jacob (2002);
Consequences	Carnoy and Loeb (2002); Carnoy (2001); Deere and Straver (2001a, 2001b)
Response to Public Disclosure	Hanushek and Rivkin (2003); Carnoy (2001)
Measurement Errors	
Testing Effects	Koretz and Barron (1998); Jacob (2002); Deere and Strayer (2001b)
Random Errors	Kane and Staiger (2001)
Exclusions/Selectivity	
	Jacob (2002); Figlio and Getzler (2002); Haney (2000); Cullen and Reback (2002); Toenjes et al. (2000); Carnoy, Loeb, and Smith (2001); Deere and Strayer (2001a, 2001b); Koretz and Barron (1998)
Other Responses	
Teacher Assignment	Boyd et al. (2002)
Achievement-Gain Accountability Systems	
Outcome Effects	
Direct Response to	Richards and Sheu (1992); Ladd (1999)
Consequences	

Table 7										
Distribution	of	Studies	of	the	Im	oacts	of	Accour	ntabil	ity

mance. Thus, as identified in Table 7, several studies indicate that exclusions from the testing tend to increase with the introduction of new accountability systems. None, however, says anything about reactions after the initial response. In most cases, the incentives for these types of reactions will decline over time.

Much less information is available about the range and scope of reactions to improve performance. In most cases studied, the introduction of a performance system has led to achievement improvements. Moreover, the response not surprisingly is more concentrated on the aspects of learning that are measured and assessed as opposed to those that are not. While some people find this to be a negative aspect of the accountability systems, it seems to be just what one would expect. The magnitude of such improvements is nonetheless not easy to characterize. Further, the exact source of the response—whether emanating from the informational aspects of the systems or from the direct sanctions and rewards—is uncertain in states where both mechanisms operate simultaneously.

Important for design considerations, information about the compar-

ative effects of alternative systems is quite limited. Understanding the differences among accountability systems requires comparing states that employ alternative approaches. It is, however, very difficult to do this. For example, Grissmer et al. (2000) interpret estimates of the superior performance of Texas and North Carolina schools on the National Assessment of Educational Progress (NAEP) as resulting from their accountability systems, but no attempt is made to test such a hypothesis formally (compare with Hanushek 2001). Carnoy and Loeb (2002) find that accountability systems that have implications for students and schools ("strong accountability") had faster growth in NAEP math achievement. Moreover, this happens not just for low-achievement students but also for high-achievement students. Nonetheless, their categorization cuts accountability systems in different ways than that previously presented. Since a number of states will soon be adopting new systems as a result of federal legislation, it is important to know which accountability features and designs produce the greatest impact on student performance measures. Specifically, it will become increasingly pertinent to know whether more costly and less understandable systems that focus on value-added measurement are significantly better than status models.

# NEW EVIDENCE ON THE IMPACT OF ACCOUNTABILITY

Inferring the impact of accountability systems is difficult both because of the recentness of their introduction in many states and because of the limited information about student performance across different accountability regimes. One source of information on performance, however, offers some possibility for analysis. NAEP has provided performance information for states during the 1990s. These examinations in mathematics track performance across grades. We use these performance measures to assess the impacts of state accountability systems. In this regard, the analysis is directly related to the work of Carnoy and Loeb (2002). It differs largely by looking at longer periods of achievement growth and by employing different measures of accountability. We also investigate whether accountability systems affect special-education placement rates by state.

### **Impacts on Student Achievement**

Understanding the impacts of different state policies on performance is difficult, in part because of the paucity of previous work describing the elements of state policy that are important. Education is the responsibility of state governments, and states have gone in a variety of directions in the regulation, funding, and operation of their schools. As a result, it is

Distribution of States by Length of	of Time with Accountability System	m, 1996 and 2000
Years with an Accountability System	1996	2000
0	41	13
1	4	10
2	2	8
3	4	6
4	0	4
5	0	4
6	0	2
7	0	4
Source: Authors' calculations.		

Table 8

difficult to assess the impacts of individual policies without dealing with the potential impacts of coincidental policy differences.<sup>12</sup>

The basic estimation approach focuses on growth of student achievement across grades. If the impacts of stable state policies enhance or detract from the educational process in a consistent manner across grades, concentrating on achievement growth implicitly allows for stable state policy influences and permits analysis of the introduction of new state accountability policies.

The NAEP testing measured math performance of fourth graders in 1992 and 1996, and of eighth graders four years after each of these assessments. While the students are not matched, the common cohort acts to eliminate a variety of common achievement influences. Our analysis of achievement relies on growth in achievement between fourth and eighth graders over the relevant four-year period (for example, growth in achievement from fourth grade in 1996 to eighth grade in 2000).13 Understanding the effect of accountability systems is dependent on the introduction of these systems. Table 8 describes the time path of introduction of accountability systems across states by reference to the length of time that accountability systems have been operating in different states. By looking at accountability systems in 1996, it is clear that much of the movement to accountability is very recent. By 1996, just 10 states had

<sup>&</sup>lt;sup>12</sup> Hanushek, Rivkin, and Taylor (1996) discuss the relationship between model specification and the use of aggregate state data. The development here builds on the prior estimation in Hanushek and Somers (2001), and the details of the model specification and estimation can be found there.

<sup>&</sup>lt;sup>13</sup> We actually rely on differences in logarithms of scores because these implicitly allow state factors to have a multiplicative effect on achievement inputs.

Table 9

	Pooled	1: 1992–96	and 1996-	1996–2000			
	(1)	(2)	(3)	(4) With state effects	(5)	(6)	(7)
Accountability or							
Report-Card System	0.0084 (3.07)	0.0096 (2.94)	0.0100 (2.56)	0.0089 (1.42)	0.0116 (2.83)	0.0131 (2.96)	0.011 (2.18)
		-0.0042				-0.0057	
Reporting System		(-1.05)				(-1.25)	
Time System in Place			-0.0006 (-0.47)				
Education of Population aged 25–29							0.0006 (0.04)
Real Spending per Pupil							0.0002 (0.03)

### Relationship of Presence of Accountability System to Improvements in NAEP Mathematics Performance

Note: All pooled estimates include an indicator variable for time period. Robust t-statistics are presented in parentheses below each coefficient. The dependent variable is log (Achievement<sub>grade 8, t</sub>/Achievement<sub>grade 4, t-4</sub>). Source: Authors' calculations as described in the text.

active accountability systems, while by 2000, just 13 states had yet to introduce active systems.<sup>14</sup>

The estimation takes two different modeling approaches to understanding the interaction of accountability systems and achievement. First, the two periods of growth between fourth and eighth grades for the states (1992 through 1996 and 1996 through 2000) are pooled, at times with extraction of state fixed effects. Second, just the latter period is used to look at cross-sectional differences in growth. The former modeling strategy is appropriate if other influences on achievement—both policy and other—are roughly constant over the entire period. The latter concentrates on the period of most activity in accountability but relies on the growth formulation with possible explicit measures of state differences to isolate the effects of accountability systems.

Table 9 presents the basic estimates of the effects of accountability systems on growth in student achievement. The simplest version (columns 1 and 5) looks at whether the state has some form of accountability system in place during the period of observation. Recall that accountabil-

<sup>&</sup>lt;sup>14</sup> In all our analyses, the universe includes 50 states plus the District of Columbia.

ity in the United States has taken two general forms—report cards and rewards/sanctions. Report cards serve a public information function whereas rewards and sanctions subject schools to material consequences. The results indicate that the presence of some form of accountability— either report cards or systems with sanctions—produce growth in achievement that is 1 percent higher than it would be without such programs. This is a large effect since the standard deviation of growth in state scores between fourth grade in 1996 and eighth grade in 2000 is just 1.2 percent.<sup>15</sup>

The remaining columns provide additional detail. The second and sixth columns show the implications of having a simple reporting system that either does not have sanctions and rewards or does not summarize the relevant performance of the school. Since reporting systems are less stringent than full accountability systems, one would expect less effect on student achievement growth. Indeed, states with reporting systems (0.42 percent versus 0.96 percent in the pooled sample), although the difference is not statistically significant. Put another way, the results show that the use of sanctions and rewards does not create a significant positive effect over the use of report cards.

With the small number of state observations it is difficult to distinguish between "no effect" and "weak data" such that precise estimation is not possible. Additionally, according to column 3, the time that the system is in effect does not appear to affect performance (that is, achievement growth moves to a higher level once the system is in place but does not continue to improve). The estimate of the overall effect of the use of accountability systems also holds even in the case of state fixed effects (column 4). Finally, while the point estimates are slightly larger when estimated just on the most recent period of achievement observation, the impact of accountability systems is virtually unchanged from that estimated by pooling the results.<sup>16</sup>

The summary of estimated effects of introducing an accountability system is simple: Accountability systems appear to lead to significantly higher growth in achievement. Of course, as discussed above, it would be nice to know more about how variations in the systems employed affect achievement. Unfortunately, the data are rather thin—fewer than 40

<sup>&</sup>lt;sup>15</sup> In all cases the dependent variable is the log of achievement growth. The introduction of an accountability system is a change from 0 to 1, which in the pooled sample corresponds to a proportional increase of 0.008, or roughly 1 percent.

<sup>&</sup>lt;sup>16</sup> Note that the last column provides estimates of achievement growth where other contemporaneous measures of state differences are included—education level and school spending per pupil of the population aged 25 to 29 (as a measure of parental education). Neither of these traditional measures of school inputs has an impact on growth in test scores, and the estimates of the effects of accountability are essentially unchanged by their inclusion.

states have complete information about achievement growth for the entire period—so it is not possible to say with any certainty whether differences in the accountability systems are important or how important they might be.

### **Special-Education Placement**

As we discussed, there is an immediate incentive in most existing accountability systems to exclude students who might be expected to have low achievement. A method often discussed is to place students into special education and thereby exclude them from testing and from subsequent inclusion in the accountability system. The previously discussed literature provides evidence from individual states and school systems suggesting that schools tend to respond in such a manner.

In order to test the importance of this incentive, we study the responsiveness of special-education placement rates to the introduction of an accountability system. We concentrate on the period 1995–2000, when a majority of the accountability systems was introduced. As with achievement analysis, our basic strategy is to relate (logarithms of) special-education placement rates to accountability and other factors that might affect placement. Unlike achievement, however, we have regular measurement of special-education placement, so that we can consider more refined models of the annual patterns in placement. It is also easy in this case to remove state differences in average special education placement (that is, state fixed effects).

Table 10 shows that the introduction of an accountability or reportcard system is associated with roughly 1.5 percentage point higher special-education placement rates in a state. These estimates are essentially generalizations of difference-in-difference estimators that allow for comparisons across all of the states. The second column indicates that the reaction to accountability occurs over time, with a 1.1 percentage point higher placement rate with accountability or report cards, and with an increase of 0.4 percentage point increase each year that the system is in place. Thus, the state estimates appear to confirm the estimates from individual states and districts.

The final three columns, however, show a markedly different picture. Specifically, throughout the nation, special-education placement rates have increased over time, and the standard methodology of comparing rates before and after introduction of accountability tends to attribute these overall increases to an effect of accountability systems. Thus, the final columns introduce a time trend and its square to allow for the strong and ubiquitous increases in special-education placement. Columns 4 and 5 show that both the effect of having an accountability or report-card system and the effect of how long such a system has been in effect have an insignificant impact on placement rates (in terms of magnitude and of

	Standard	Approach	Allowar	Allowance for Placem		
	(1)	(2)	(3)	(4)	(5)	
Accountability or Report Card System	1.45 (10.1)	1.09 (7.9)	.11 (1.0)	.10 (.9)	.09 (.7)	
Time in Place		.38 (7.9)		02 (5)		
Time Trend			.86 (12.4)	.87 (14.4)	.87 (12.5)	
Time Trend Squared			08 (-6.3)	08 (-6.0)	08 (-6.4)	
Report Card System					.24 (1.2)	
Longitudinal System					73 (-1.9)	

#### Table 10

Effect of Accountability on Special-Education Placement Rate, 1995 through 2000

Note: Estimation employs a panel of special-education placement rates for all states and the District of Columbia over the period 1995–2000. Estimation includes a fixed effect for each state. The t-statistics appear in parentheses below each estimate.

Source: Authors' calculations as described in the text.

statistical significance). The final column introduces the characteristics of the state system. Report card states seem to have a slight positive influence on placement rates. Longitudinal accountability systems (the cohort-change and individual-gain approaches used in several states) lower placement rates, perhaps reflecting regulations on accountability along with the incentives discussed earlier. While neither of these estimates is statistically significant, the impact of longitudinal systems is close to standard levels (p>0.06)—even though there are very few observations of such systems.

These estimates suggest caution in interpreting analyses of the gaming of accountability systems. If such gaming were generally important, it should show up in the national data—but it does not. Moreover, the national trends in special-education placement offer a ready explanation for the divergent results.

### Some Conclusions

One of the major conclusions to be drawn from this discussion is that the existing body of evidence about accountability systems is fairly sparse. Moreover, much of it does not help to diagnose the various sources of incentive impacts. Without greater attention across states to understanding the "signal-to-noise" characteristics of the systems in place, policymakers run the risk of confounding the true effects of their efforts with factors outside of their control.

The analysis provides some simple but powerful messages about state accountability systems. To begin with, on a conceptual level most of the existing systems that have been introduced are not good devices for inferring the quality of individual schools. As a result, they are also not good devices for providing incentives. The incentives do not accurately relate to the activities and performance of the schools, and they are subject to a variety of approaches to "game" the system. These design problems may reflect not having thought out the issues; alternatively they may reflect simple politics that hamstring the introduction of better incentive systems.

The design problems occur in a variety of different forms. Some systems confuse student performance with the inputs and behavior of the schools. Other systems make it difficult if not impossible to separate effects on outcomes that are related to school performance from effects of parents or past educational inputs.

A review of the extant information on how schools react to accountability systems suggests that schools do indeed react to the introduction of accountability systems. At the same time, not all of the reactions appear to be desirable. A variety of investigations of attempts of schools to alter measured achievement without necessarily changing the reality indicates that schools do operate on this margin. Nonetheless, while discovering such unintended consequences is good sport for academics, one would expect the immediate gaming to be much more important than any continual gaming. In other words, this kind of behavior appears largely self-correcting.

Most of the initial investigations also show that the introduction of accountability systems leads states to improve on performance. The confusion with artificial increases through gaming or with responses tailored very specifically to the state testing, however, makes the evidence a little difficult to interpret.

In order to dig more deeply into the effects of accountability systems, we have conducted two new analyses of accountability in the states. We look across the states and investigate whether the introduction of accountability is associated with greater growth in achievement and whether it is associated with more placements into special education. On the first score, we find that achievement growth between the fourth and the eighth grade is 1 percent higher after the introduction of a state accountability system. Further, the differences in impact on achievement between the use of report cards (public disclosure of performance data) and systems that expose schools to direct consequences based on scores are not significant, suggesting that the "power" of accountability lies in reducing barriers to information rather than rewards or punitive measures. The data are not good enough, however, to give us much

confidence in whether or not different types of systems have a differential effect.

On the latter score, we find that special-education placement does not appear closely related to the introduction of accountability in a state. Special-education placement rates have increased over time. Once this is allowed for, the introduction of an accountability or report card system has no significant impact on special-education placement, suggesting some caution in interpreting the prior evidence for longitudinal changes within states or districts.

An important element of this analysis is simply setting out some of the features that we believe are most important in thinking about accountability. Specifically, most existing systems—when seen from the perspective of incentives for schools—are seriously flawed. At the same time, we know that they have an ability to evoke responses from schools. It would be most unfortunate if we lumped all accountability systems together and concluded on the basis of our early observations that they lead to some bad outcomes and thus should be eliminated. This is simply not the message that should be taken from the existing reactions.

If we are interested in student achievement—as we should be—we simply have to focus on student achievement. This is the genius of accountability systems. The perspective should not be whether or not to eliminate accountability but instead how to refine it to provide the kinds of incentives that we want.

Perhaps more important, because accountability is often viewed as a binary choice—you either have it or you don't—it is very likely that some, or even most, of the existing systems will not stand up to expectations. It would be inappropriate, however, to conclude that greater accountability does not work on the basis of results from most existing state systems.

### References

- Baker, George P. 1992. "Incentive Contracts and Performance Measurement." Journal of Political Economy 100 (3): 598-614.
- —. 2002. "Distortion and Risk in Optimal Incentive Contracts." Journal of Human Resources 37 (4): 724-51.
- Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." Quarterly Journal of Economics 114 (2): 577-99.
- Boyd, Don, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2002. "Do High-Stakes Tests Affect Teachers' Exit and Transfer Decisions? The Case of the 4th Grade Test in New York State." Stanford Graduate School of Education, working paper (July).
- Brownson, Amanda. 2001. "Appendix B: A Replication of Jay Greene's Voucher Effect Study Using Texas Performance Data." In *School Vouchers: Examining the Evidence*, edited by Martin Carnoy. Washington, DC: Economic Policy Institute.
- Carnoy, Martin. 2001. School Vouchers: Examining the Évidence. Washington, DC: Economic Policy Institute.
- Carnoy, Martin and Susanna Loeb. 2002. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." Stanford Graduate School of Education, unpublished paper (March).
- Carnoy, Martin, Susanna Loeb, and Tiffany L. Smith. 2001. "Do Higher State Test Scores in

Texas Make for Better High School Outcomes?" Paper presented at the American Educational Research Association Annual Meeting (April).

- CREDO. 2002. "The Future of California's Academic Performance Index." CREDO, Hoover Institution, Stanford University (April).
- Cullen, Julie B. and Randall Reback. 2002. "Tinkering Toward Accolades: School Gaming under a Performance Accountability System." Department of Economics, University of Michigan, working paper.
- Deere, Donald and Wayne Strayer. 2001a. "Closing the Gap: School Incentives and Minority Test Scores in Texas." Department of Economics, Texas A&M University, working paper (September).
  - . 2001b. "Putting Schools to the Test: School Accountability, Incentives, and Behavior." Private Enterprise Research Center Working Paper No. 0113. Texas A&M University (March).
- Dixit, Avinash. 2002. "Incentives and Organizations in the Public Sector: an Interpretative Review." *Journal of Human Resources* 37 (4): 696-727.
- Figlio, David N. and Lawrence S. Getzler. 2002. "Accountability, Ability and Disability: Gaming the System" NBER Working Paper No. 9307 (November).
- Figlio, David N. and Maurice E. Lucas. 2000. "What's in a Grade? School Report Cards and House Prices." NBER Working Paper No. 8019 (November).
- Gramlich, Edward M. and Patricia P. Koshel. 1975. *Educational Performance Contracting*. Washington, DC: Brookings Institution.
- Greene, Jay P. 2001a. "An Evaluation of the Florida A-Plus Accountability and School Choice Program." Center for Civic Innovation, Manhattan Institute (February).
  - 2001b. "The Looming Shadow: Florida Gets Its 'F' Schools to Shape Up." *Education Next* 1 (4): 76-82.
- Grissmer, David W., Ann Flanagan, Jennifer Kawata, and Stephanie Williamson. 2000. Improving Student Achievement: What NAEP State Test Scores Tell Us. Santa Monica, CA: Rand Corporation.
- Haney, Walter. 2000. "The Myth of the Texas Miracle in Education." Education Policy Analysis Archives 8 (41) <epaa.asu.edu>.
- Hanushek, Eric A. 2001. "Deconstructing RAND." Education Matters 1 (1) Spring: 65-70.
- Hanushek, Eric A. and Margaret E. Raymond. 2001. "The Confusing World of Educational Accountability." *National Tax Journal* 54 (2): 365-84.
- Hanushek, Eric A. and Steven G. Rivkin. 2003. "Does Public School Competition Affect Teacher Quality?" In *The Economics of School Choice*, edited by C. M. Hoxby. Chicago: University of Chicago Press.
- Hanushek, Eric A. and Julie A. Somers. 2001. "Schooling, Inequality, and the Impact of Government." In *The Causes and Consequences of Increasing Inequality*, edited by F. Welch. Chicago: University of Chicago Press.
- Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2001. "Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools." NBER Working Paper No. 8479 (September).
- ———. Forthcoming. "Inferring Program Effects for Specialized Populations: Does Special Education Raise Achievement for Students with Disabilities?" *Review of Economics and Statistics*.
- 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 (4): 611-27.
- Jacob, Brian A. 2002. "Making the Grade: The Impact of Test-Based Accountability in Schools." Kennedy School of Government, Harvard University, working paper (April).
- Kane, Thomas J. and Douglas O. Staiger. 2001. "Improving School Accountability Measures." NBER Working Paper No. 8156 (March).
- Koretz, Daniel M. and Sheila I. Barron. 1998. The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS). Santa Monica, CA: RAND Corporation.
- Ladd, Helen F. 1999. "The Dallas School Accountability and Incentive Program: An Evaluation of the Impacts of Student Outcomes." *Economics of Education Review* 19 (1): 1-16.
- Ladd, Helen F. and Elizabeth J. Glennie. 2001. "Appendix C: A Replication of Jay Greene's Voucher Effect Study Using North Carolina Data." In *School Vouchers: Examining the Evidence*, edited by M. Carnoy. Washington, DC: Economic Policy Institute.

- Richards, Craig E. and Tian Ming Sheu. 1992. "The South Carolina School Incentive Reward Program: A Policy Analysis." *Economics of Education Review* 11 (1): 71-86.
- Rose, Lowell C. and Alec M. Gallup. 2001. "The 33rd Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes toward the Public Schools." *Phi Delta Kappan* 83 (1): 41-58.
- Sanders, William L. and Sandra P. Horn. 1994. "The Tennessee Value-Added Assessment System (TVAAS): Mixed-Model Methodology in Educational Assessment." Journal of Personnel Evaluation in Education 8 (3): 299-311.
- Toenjes, Laurence A. and A. Gary Dworkin. 2002. "Are Increasing Test Scores in Texas Really a Myth, or Is Haney's Myth a Myth?" *Education Policy Analysis Archives* 10 (17) <epaa.asu.edu>.
- Toenjes, Laurence, A., A. Gary Dworkin, J. Lorence, and A.N. Hill. 2000. "The Lone Star Gamble: High Stakes Testing, Accountability and Student Achievement in Texas and Houston." Sociology of Education Research Group, University of Houston, working paper (August).
- Weimer, David L. and Michael J. Wolkoff. 2001. "School Performance and Housing Values: Using Non-Contiguous District and Incorporation Boundaries to Identify School Effects." National Tax Journal 54 (2): 231-53.

# Discussion

# IMPROVING EDUCATIONAL QUALITY: How Best to Evaluate Our Schools?

# Peter J. Dolton\*

Hanushek and Raymond have provided us with an elegant and coherent set of arguments for accountability in the public provision of education. Their main empirical contribution is the characterization of the different types of accountability systems in the United States. They go on to show how the operation of incentives in the public sector is not easy to administer and evaluate. Specifically they draw attention to the distinction between inputs, process variables, and outcomes, and they show how incentive systems reliant on input and process measurements may be ineffective. The authors find empirically that the presence of an accountability system leads to modest growth in achievement but caution that evidence needs to be treated carefully to recognize the possibility of gaming and the consequent interpretation problems. In short, they suggest that accountability incentives matter.

This discussion of their paper recaps the problems with incentive structures in public provision of education and raises the issue of exactly what is meant by accountability. I also consider how the effect of accountability on achievement measurement and on performance should be assessed. Then I will provide examples of the problematic working of incentives from the U.K. education system.

# WHAT ARE THE PROBLEMS WITH INCENTIVE STRUCTURES?

There is now widespread evidence that incentives work in the public sector. The issue is designing incentive structures that are not subject to distortion or "gaming." The education production process is very reliant

<sup>\*</sup>Professor of Economics, University of Newcastle and University of London.

on teacher labor as the most important factor of production. In practice, it is very difficult to write complete labor contracts in education to generate the appropriate incentives for teachers. To a large extent this is a principal/agent problem. However there are several extra dimensions to this problem in education.

The literature (Dixit 2002 and Burgess and Metcalfe 1999) suggests that using incentive structures in the public sector could induce dysfunctional behavior in the sense that employees could direct their effort on some aspects of their work, to the detriment of other aspects, or in a counterproductive way, when teamwork or the cooperation of colleagues is involved.

The essential problem of public sector educational provision is that education is not a single output, and any education system must have multiple goals. Dixit (2002) lists the multiple goals of public education as the following:

- 1. Imparting basic skills of literacy, mathematics, and science for communication, reasoning, and calculation;
- 2. Fostering the emotional and physical growth of children;
- 3. Preparing students for work, by teaching them vocational skills and attitudes suitable for employment;
- 4. Preparing them for life, by teaching them skills of health and financial management;
- 5. Preparing them for society, by instilling ideals of citizenship and responsibility;
- 6. Helping them to overcome disadvantageous circumstances at home, including in many cases poor nutrition and poor study environments; and
- 7. Providing an environment free from drugs and violence.

Dixit suggests that although these goals are not mutually contradictory, they do compete for resources. To this degree they are alternative outputs in the educational production process, and teacher effort put into one of these objectives may detract wholly, or in part, from one or more of the other goals. Hanushek and Raymond view educational production in much narrower terms, as very few of these goals appear on their list of accountability variables.

The essential problem of education is that with multiple goals, it is unclear how to direct effort. Holmstrom and Milgrom (1991) develop a model that explains that the way incentives work may not be appropriate, even when accurate performance measures are available. They extend the standard principal/agent model to one in which there are several dimensions to effort. The general result is that the agent will have an incentive to divert effort away from the less accurately measured task. Hence, it is shown that if the principal wishes the agent to allocate effort towards a task that is not easily measured, then incentives on the measurable tasks must be weakened.

The second essential feature of any education system is that it has multiple principals. As a consequence, the actions of any individual teacher (agent) could be affected by many other people (principals) who are in a position of influence. Most specifically the wishes of parents, headteachers, teacher unions, local or federal authorities, taxpayers, employers, religious and ethnic pressure groups, governors, and even pupils may influence the actions and decisions of individual teachers. However, Dixit (1997) shows (under regularity conditions) that the existence of several principals makes the overall incentives for the agent much weaker. This weakening of incentives occurs because each principal will seek to divert the agent's effort to his most preferred dimension. Obviously the more principals that are involved with competing interests the more diluted will be the incentive structure for the agent. Hanushek and Raymond do not discuss the multiple principal incentive problem or its implications for accountability.

# WHAT DO WE MEAN BY ACCOUNTABILITY?

The concept of accountability is a difficult one. Fearon (1999) suggests that "one person, A (the agent), is accountable to another B (the principal), if two conditions are met. First, there is an understanding that A is obliged to act in some way on behalf of B. Second, B is empowered by some formal institution or perhaps informal rules to sanction or reward A for her activities or performance in this capacity" (p. 55).

Laver and Shepsle (1999) provide a different definition. They suggest that a political agent is accountable to a principal when the principal, having the means to do so, has no inclination to replace the agent with a feasible alternative. Hence Laver and Shepsle view accountability as both an equilibrium state and a mechanism for change.

Ferejohn (1999) suggests there are three serious limits to accountability. First, the nature of the accountability mechanism (voting rule) may mean that minorities are ignored or indeed that electoral heterogeneity makes it possible for officials to play off some voters against others. Second, the institutions of accountability operate in real time—and this provides the officials with the opportunity to avoid responsibility. Third, officials typically enjoy an immense informational advantage over consumers.

Hanushek and Raymond do not have a clear definition of what they mean by accountability. They suggest that accountability creates incentives—although they recognize that such incentives may not always have desirable consequences. What is unclear in their exposition is whether they believe the mere publication of information on standards in schools will provide an adequate incentive for efficient resource allocation. Surely a necessary (but not sufficient condition) for such efficiency is that this accountability be directly linked to a quasi market, or the power of consumers to choose alternative providers in a competitive market.

Hence what I am suggesting is that effective accountability in education necessitates, first, that the education system provide consumers with full information to make decisions; second, that consumers have the power to influence the balance of priorities across the multiple goals of educational provision; third, that consumers have the means to choose alternative providers in a competitive or quasi-competitive environment; and, fourth, that any incentives that operate on education providers do not act to distort their incentives regarding their provision in ways that are counter to the wishes of consumers.

A definition of accountability that includes customers' wishes relies on being able to identify who these customers are; identifying correctly what their views are; aggregating their views into a consensus to establish what the ranking of priorities is; and then implementing these views effectively. Any decisions concerning public expenditure and investment in education constitute a social-choice problem (see Majumdar 1983). This issue is rarely examined.

We also have to assume that the views of parents are responsible and representative of the whole customer base. Such an assumption may be unrealistically ideal. Aoki and Feiner (1996) discuss how the parents whose views are more effectively heard are disproportionately those who live in affluent areas, are more highly educated, and have higher-status occupations. Such evidence means that establishing precisely who an educational system is accountable to, and what the mechanism is for the transmission of the influence, is important. Most concretely, are the customers of education the parents or the pupils? Undoubtedly the priorities of the pupils, if consulted, may be different from their parents'.

At the heart of effective public service provision is the possibility of competition among providers. Unless there are alternative schools for parents to send their children to, there is no incentive mechanism for each school to compete in the quasi market. Another problem with this model arises if there are private schools outside the public sector. Friedman (1962) advocates a voucher system in which essentially all schools would be private. In the context of the present system where private and state schools operate in the same area, there is the "exit and voice" issue, which states that there will not be an effective mechanism for change if the most influential parents choose to "exit" from the state schools to the private schools. More research is necessary into how the public sector in education can efficiently co-exist alongside a private sector (see Hirschman 1970).

## HOW CAN EDUCATIONAL OUTPUT AND EFFICIENCY BE **MEASURED? WHAT IS THE EFFECT OF ACCOUNTABILITY?**

Hanushek and Raymond show that most states rely on student performance as the main outcome measure. They come to the conclusion that the "individual-gain score" (or value-added) measure of student performance between years is the most valid. It is worth reviewing exactly what such measures can and cannot say. Todd and Wolpin (2003) provide the most general discussion of the assessment of educational evaluations and rigorously discuss modeling using the value-added method. To fix ideas, we consider that pupil attainment is determined by a production function relation, and we may use the following notation:

- $A_{ijkt}$  attainment of pupil *i*, in class *j*, in school *k*, at the end of time period  $t^1$
- characteristics of pupil *i* at time *t* that may affect attainment  $X_{it}$
- $S_{ijkt}^{''}$  resources of pupil *i*'s class *j*, in school *k*, at time *t*  $F_{it}$  family resources devoted to pupil *i* at time *t*
- innate ability endowment of pupil *i*  $\mu_i$

Assuming that educational attainment of the pupil is a function of individual attributes and ability, school inputs, and family inputs, we may write the general production function type model of what determines pupil attainment as:

$$A_{ijkt} = g(X_{it}, S_{ijkt}, \Sigma F_{it}, \mu_i).$$
<sup>(1)</sup>

Simplifying this production relation to consider the influence on initial attainment, prior to school, we suggest that

$$A_0 = g_0(X_0, F_0, \mu),$$
 (2)

where we are dropping the i subscript for the individual. In each subsequent period the family adds more input based on its decision process, and the school contributes resources, S, in the manner suggested by the production function. The schooling input decision for any pupil will be determined as a result of the pupil's ability and prior attainment, that is,

$$S_{ijkt} = \varphi(A_{ijkt-1}, \mu_i). \tag{3}$$

We can write the production function (1) as an econometric model for period 1 as:

<sup>&</sup>lt;sup>1</sup> For notational convenience we will think of pupils' attainment being tested at the end of each school year t, so that A<sub>i</sub> represents the attainment acquired in that year.

$$A_{ijk1} = \alpha_1 X_{i1} + \beta_1 S_{ijk1} + \delta_1 F_{i1} + \gamma_1 \mu_i + \epsilon_{i1} + u_{j1}, \tag{4}$$

where  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$  are parameters and  $\epsilon_i$  and  $u_j$  represent unobserved heterogeneity at the individual and school level. Likewise for period 2 we can write,

$$A_{ijk2} = \alpha_2 X_{i2} + \beta_2 S_{ijk2} + \delta_2 F_{i2} + \gamma_2 \mu_i + \epsilon_{i2} + u_{j2}.$$
 (5)

If we take the difference between (5) and (4), we get an expression for the fixed effects estimator that is so common in econometric applications. Writing this as:

$$\Delta A \equiv A_{ijk2} - A_{ijk1},\tag{6}$$

we can see that this is equivalent to

$$\Delta A = \alpha_2 X_{i2} - \alpha_1 X_{i1} + \beta_2 S_{ijk2} - \beta_1 S_{ijk1} + \delta_2 F_{i2} - \delta_1 F_{i1} + \gamma_2 \mu_i - \gamma_1 \mu_i + \epsilon_{i2} - \epsilon_{i1} + u_{j2} - u_{j1}$$
(7)

The question is, under what circumstances is the individual student-gain estimator a valid estimate of pupil progress? We place some restrictions on this model to make explicit the necessary assumptions:<sup>2</sup>

- A1. The pupil attributes,  $X_i$ , remain constant across time. This means we can write:  $\alpha_2 X_{i2} \alpha_1 X_{i1} = \alpha X_i$ , where  $(\alpha_2 \alpha_1) = \alpha$  and  $X_{i1} = X_{i2}$ . This is a restrictive assumption since it means that variables that represent motivation and effort, like propensity to complete homework, remain fixed. This is clearly wrong, as such attributes are often age-related for the pupil.
- A2. There exists a sufficient statistic for the changing value of school inputs that is observable and that school effects are time invariant. This means we can write  $S_2 S_1 = S$ . Then we can write the school effects term as  $\beta S$ .
- A3. The change in parental input can be proxied by some observable family characteristic *F*. This is equivalent to assuming that  $F_2 = F_1$  and  $\delta_2 \delta_1 = \delta$  and hence the family effects can be represented by  $\delta F$ . Although naive and restrictive, it is unlikely that values of family inputs will be observed at different points in time.

 $<sup>^{\</sup>rm 2}$  It should be appreciated that more than one set of assumptions can be made in order to make this model useful with the data.

- A4. The impact of ability endowment on achievement is independent of time. Hence  $\gamma_2 = \gamma_1$  and so the unobservable ability term can be netted out. This would suggest that the importance of the application of innate ability is not age-specific. Again, this is naïve, as maturity may well affect the pupil's potential for attainment.
- A5. Input choices made by schools and parents are invariant to prior achievement outcomes. Thus  $S_1$  and  $F_1$  are uncorrelated with  $A_0$ . Taking A1 through A5 together we can now rewrite (7) to give:

$$\Delta A = \alpha X_i + \beta S_{ijk} + \delta F_i + \epsilon_i + u_j. \tag{8}$$

As restrictive as the above assumptions are, the possible estimation of equation (8) still has very demanding data requirements.

The problem of Hanushek and Raymond is more complex. They wish to establish the effect of introducing accountability on student achievement. One approach would be to estimate equation (8)—possibly aggregated at the level of the school or the state, splitting the sample according to whether the state operated with accountability. This would require that the decision to invoke accountability was exogenous to achievement. Their adopted approach, due to inadequacies in their data, is less sophisticated. They use state-level data with no information on school or family resource decisions. They use their own measure of accountability as a regressor into equation (8). The question is, under what circumstances is it valid to assume that such a regressor is exogenous with an additively separable effect? Clearly one would expect the level of school resources and family factors to be affected by the level of accountability.

There are other subtle ways in which the inputs and outputs of the education production process are difficult to observe. The raw material, or input, a teacher works with is highly variable. It is well known that teaching the same material to children from poor homes in deprived areas is more difficult than teaching to motivated children from middleclass homes. Even if one tries to measure value-added in terms of improvement of exam scores, these can be a distortion of the improvement in attainment as such a calculation assumes that other factors and their influence are fixed over time. There is often a huge variation in the resources at the school's disposal—many of which are not easily measured.

A second often-overlooked issue is the measurement of peer effects in schools (see Lazear 2001). It is possible that some of the results relating to the absence of a pupil-teacher ratio resource effect (see Hanushek 1997 and Burtless 1996) may be due to ignoring peer effects. Indeed, one study in the United Kingdom shows this dramatically with a complete change in the resource coefficient when peer-group effects are proxied. (See Dolton 2002a.)

A third important limitation of achievement gains models is that a pupil's learning may not be apparent until years after his schooling. Often the value of what is learned by the pupil is not used or tested until several years after. As Hanushek and Raymond point out, one further important limitation to the value-added model is that achievement gains cannot be identified for those who move geographical location.

Finally, it is not impossible that teachers (or principals other than government) may view educational output differently from the government. Teachers may want to promote curiosity, induce creative thinking, provide pastoral care, and develop a wider curriculum. The government may prefer to structure the curriculum, standardize teaching methods, meet minimum standards on basic skills, and maximize performance on SAT test scores. The wider benefits of learning are rarely added into achievement-gain calculations.

# INCENTIVES AND QUASI MARKETS IN U.K. SCHOOLS

There has been a major shift in the way in which public sector education has been provided in the United Kingdom over the last 20 years. The educational system has changed to one dominated by incentive structures and quasi markets. These changes have produced a revolution in state educational provision. The results and consequences so far have been mixed. I will highlight how several of these quasi markets have been working, including some of their unintended consequences.

In the United Kingdom the Education Reform Act was passed in 1988. The general aim of the reform was to introduce a more competitive quasi-market approach to the allocation of resources in the education system. It introduced financial delegation to schools, and this involved the introduction of "formula funding" in which school income is based directly on pupil numbers. The Act insisted on the publication of school league tables and introduced the principle that parents had the right to send their children to any school they wished. The idea was that popular schools were allowed to expand without limit and conversely unpopular schools, mostly in inner cities, to contract or even close. The principles of parental choice and devolved school funding linked directly to pupil numbers establishes the conditions under which-theoretically-a quasi market can operate. This approach was designed to provide teachers and schools with appropriate incentives for efficiency and effectiveness. Although requiring schools to live within their budgets, this approach does not provide the same incentives for employees as knowing that their efforts contribute to the profit "bottom line" of a firm.

One clear feature of the state education system in the United

Kingdom is that there is a lack of competition. State schools in the United Kingdom, in many areas, operate essentially as monopoly providers. Only around 7 percent of school children in the United Kingdom attend independent schools. Because of the scale of their fees, these independent schools do not present a realistic alternative to state schools for most parents. It was this lack of competition that was part of the rationale for the 1988 Education Act providing parents with choice. The central idea behind the creation of a quasi market in state education is the theory that the introduction of competition would provide the appropriate incentives to schools to become more efficient. Theoretically this, in turn, may provide incentives for teachers to improve their performance. However, this naive faith in the power of market forces must be tempered by the reality that multiple tasks and multiple agents will weaken the power of such incentive structures.

Empirical evidence from the United States (Chubb and Moe 1990) supports the view that decentralized schooling systems produce better results, measured in terms of educational outcomes. The 1988 Act also devolved the administrative and financial control of schools to each headteacher and the school's governing body. The governing body was also to have representation from parents.

Bartlett (1993) reports that the effect of the reform has been a large shift in the distribution of resources between schools. Schools in the poorest inner-city areas have received reduced funding while funding has increased for schools in the more prosperous areas of the country. Likewise, the appointment of proactive parent governors in middle-class areas is straightforward but finding any parents willing to do the job in deprived areas is difficult. Overall, the effect of the quasi-market reforms on educational outcomes and efficiency in the United Kingdom is hard to judge, not least because there are several initiatives acting on the market at the same time. Nevertheless, there are some microeconometric studies that suggest that efficiency improvements can be directly attributed to the quasi market (Bradley, Johnes, and Millington 2001).

In reality, access to oversubscribed schools remains rationed with some selectivity and "cream-skimming" operations. This has been reflected in the market-clearing mechanism of rising house prices in localities with the best performing schools. (See Gibbons and Machin 2002).

Since 1995, the government has published school league tables of the results of all schools in the United Kingdom based on national examinations for pupils aged 7, 11, 14, 16, and 18. Some commentators, for example, Glennerster (2002), have suggested that these results show how educational standards have improved in the United Kingdom over the last six years. Table 1 shows a remarkable rise in the performance of 14-year-olds in the United Kingdom on reading, math, and science. The proportion reaching the expected standard in reading has risen from 49

United Kin Percent Rea	gdom Natio aching Expe	onal Achiev cted Levels	rement Tes	ts at Level	3, Aged 14		
	1995	1996	1997	1998	1999	2000	2001
Reading	49	57	67	71	78	83	81
Math	45	54	62	58	69	72	70
Science	70	62	69	69	79	85	87
Source: Gleni	nerster (2002)	, Table 6.					

percent in 1995 to 81 percent in 2001. In math, the proportion has risen from 45 percent to 70 percent; and science has jumped from 70 percent to 87 percent over the same period. Such statistics raise the following questions:

- To what extent are these tests based on absolute standards that have not been manipulated by a government that has declared, as if by decree, that educational standards will rise over the next five years? Alternatively, have the exams become easier or have the pupils improved their performance over time because of the predictable nature of the exams and rote learning?
- To what extent has there been misallocation of resources towards median and marginal pupils at the threshold of achievement levels in order to maximize the number of pupils passing the thresholds?
- Has the introduction of these tests diverted resources away from the least able and most able, towards the average child?
- If the improvement has been real—is it really a treatment effect that results directly from the operation of the quasi market rather than a redirection of effort on literacy and numeracy in the curriculum?
- Are the long-term consequences of increasing marginal standards on narrowly focused tests in math and English valuable for long-term educational objectives like citizenship and transferable skills?
- Is it possible to reconcile these data with results from Gundlach, Wöessman, and Gmelin (2001), who suggest that the United Kingdom along with other OECD countries has experienced a dramatic fall in school productivity over the last 25 years?

# **POSTSCRIPT: A LESSON FROM HISTORY?**

The attempt to introduce incentives and monitoring into schools in order to secure their efficiency and make the best use of public money is not new. Neither is the possibility that such attempts may lead to counterproductive consequences of these incentives.

In 1857, the Newcastle Commission surveyed schools in Great Britain and recommended "a searching examination by competent authority of every child in every school to which grants are to be paid with a view to ascertaining whether these indispensable elements of knowledge are thoroughly acquired, and to make the prospects and position of the teacher dependent to a considerable extent on the results of this examination" (Armytage 1964, p. 124).

Armytage (1964) reports on the unforeseen by-products that led to many unfortunate practices: "The cult of the 'register,' acquiescence in large classes, the deliberate cultivation of rote-memory to defeat the inspectors; even, we are told, the presentment of sick children for attendance grant." The possibility of gaming the system and its consequences were recognized even then. Matthew Arnold reported the process "as a game of mechanical contrivance in which the teachers will and must more and more learn how to beat us" (Report of the Committee of Council for 1865, p. 291, quoted in Armytage 1964, p. 125).

After 30 years, the system of public funds based on performance was abolished largely because of the problem of designing the appropriate incentives. Perhaps we can learn a lesson from history.

### References

- Armytage, W. H. G. 1964. Four Hundred Years of English Education. Cambridge: Cambridge University Press.
- Aoki, Masoto and Susan Feiner. 1996. "The Economics of Market Choice and At-Risk Students," in Assessing Educational Practices: The Contribution of Economics, edited by W. Becker and W. Baumol. New York: Russell Sage Foundation.
- Bartlett, Will. 1993. "Quasi-Markets and Educational Reform," in *Quasi-Markets and Social Policy*, edited by J. LeGrand and W. Barlett. New York: Macmillan.
- Bradley, Steve, Geraint Johnes, and Jim Millington. 2001. "School Choice, Competition, and the Efficiency of Secondary Schools in England." *European Journal of Operational Research* 135: 527–44.
- Burgess, Simon and Paul Metcalfe. 1999. "Incentives in the Public Sector: A Survey of the Evidence." CMPO Discussion Paper 99/016, University of Bristol.
- Burtless, Gary, ed. 1996. Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success. Washington, DC: Brookings Institution.
- Chubb, John E. and Terry M. Moe. 1990. *Politics, Markets and America's Schools*. Washington, DC: Brookings Institution.
- Dixit, Avinash. 1997. "Power of Incentives in Public versus Private Organizations." American Economic Review 87 (2): 378–82.

——. 2002. "Incentives and Organizations in the Public Sector: An Interpretative Review." *Journal of Human Resources* 37 (4): 696–727.

- Dolton, Peter. 2002a. "Evaluating Education Inclusion." University of Newcastle-upon-Tyne, unpublished paper.
- ———. 2002b. "Is Performance Related Pay for Teachers Possible?" Institute of Education, unpublished paper.
- Dolton, Peter, Stephen McIntosh, and Arnaud Chevalier. Forthcoming. "Teacher Pay and Performance: A Review of the Literature." *Bedford Way Papers*. London: Institute of Education Publications.
- Fearon, James. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Types Versus Sanctioning Poor Performance" in *Democracy, Accountability and Repre*-

sentation, edited by A. Przeworski, S. Stokes, and B. Manin. Cambridge: Cambridge University Press.

- Ferejohn, John. 1999. "Accountability and Authority: Toward a Theory of Political Accountability," in *Democracy, Accountability and Representation*, edited by A. Przeworski, S. Stokes, and B. Manin. Cambridge: Cambridge University Press.
- Friedman, Milton. 1962. "The Role of Government in Education," in *Capitalism and Freedom*. Chicago: University of Chicago Press.
- Gibbons, Steve and Stephen Machin. 2002. "Valuing Primary Schools." London School of Economics, unpublished paper.
- Glennerster, Howard. 1991. "Quasi-Markets for Education?" Economic Journal 101 (408): 1268-76.
- ——. 2002. "United Kingdom Education 1997–2000." Oxford Review of Economic Policy 18 (2): 120–36.
- Gundlach, Erich, Ludger Wöessman, and Jens Gmelin. 2001. "The Decline of Schooling Productivity in OECD Countries." *Economic Journal* 111 (471): C135–47.
- Hanushek, Eric. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." Educational Evaluation and Policy Analysis 19 (2): 141–64.
- Hirschman, Albert O. 1970. Exit, Voice, and Loyalty. Cambridge, MA: Harvard University Press.
- Holmstrom, Bengt and Paul Milgrom. 1991. "Multi-Tasking Principal-Agent Analyses: Linear Contracts, Asset Ownership, and Job Design." Journal of Law, Economics, and Organization 1 (7): 24–52.
- Ladd, Helen F. 1996. Holding Schools Accountable: Performance-Based Reform in Education. Washington, DC: Brookings Institution.
- Lazear, Edward. 2001. "Educational Production." *Quarterly Journal of Economics* 116 (3): 777–804.
- Laver, Michael and Kenneth Shepsle. 1999. "Government Accountability in Parliamentary Democracy," in Democracy, Accountability and Representation, edited by A. Przeworski, S. Stokes, and B. Manin. Cambridge: Cambridge University Press.
- Majumdar, Tapas. 1983. Investment in Education and Social Choice. Cambridge: Cambridge University Press.
- Przeworski, Adam, Susan Stokes, and Bernard Manin, eds. 1999. Democracy, Accountability and Representation. Cambridge: Cambridge University Press.
- Todd, Petra and Kenneth Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal* 113 (485): F3–33.

# Discussion

# IMPROVING EDUCATIONAL QUALITY: How Best to Evaluate Our Schools?

# Thomas J. Kane\*

While the academic debate has been preoccupied for much of the last decade with school vouchers, state policymakers have been moving in a very different direction, constructing elaborate incentive systems using school-level test-score measures. For instance, California spent nearly \$700 million on school-level incentives in 2001, providing bonuses of up to \$25,000 per teacher in schools with the largest increases in test performance between 1999 and 2000. Unfortunately, given that the discipline of economics has a long tradition of thinking about the design of incentives, economists have been largely absent from the debate accompanying the design of school accountability systems. For anyone seeking to catch up with the policy debate, the Hanushek and Raymond paper is extremely useful in categorizing the types of systems that have been created, in summarizing the fledgling literature on the impact of school accountability systems within states, and in providing some original evidence on the impact of state accountability policies using the National Assessment of Educational Progress (NAEP) test scores across states over time.

One of the great contributions of the paper is that it simply provides a clearer picture of the variety of systems that have been created. As described by the authors, most systems use a hybrid of one of three types of measures of test performance: status measures (mean levels of test performance), status-change measures (changes in the level of performance between cohorts over time), and gain score (the mean improvement in performance for a given cohort of students). Possibly because of

<sup>\*</sup> Professor of Policy Studies and Economics at the University of California, Los Angeles.

the difficulty of tracking individual students' performance over time, most states have chosen to base their systems on either status measures or status-change measures.

# LESSONS FROM OPTIMAL INCENTIVES LITERATURE IN ECONOMICS

A number of potential lessons can be learned from the optimal incentives literature in economics. First, as is hinted by the authors, incentive systems based upon status-change measures inevitably are subject to "ratchet effects." Raising the bar in the future based upon performance today forces schools to choose between the payoff of improvements today and the increased cost of maintaining that level of performance in the future. It is particularly striking when Hanushek and Raymond point out that evaluations based on changes in performance are a component in most state accountability systems. When performance today has an impact on expectations tomorrow, schools may underinvest in reform. (This is particularly true in systems measuring status change for single grade levels, since those using multiple grade levels may continue to benefit from any pedagogical improvement for several years as a given cohort of students moves through several grade levels.) The "ratcheting" problem is exacerbated by the fact that rewards are usually discontinuous, stair-shaped functions of performance-meaning that the magnitude of one's reward is not a function of the distance by which a school might clear a given threshold. The authors note that in a system based upon status changes, a school may generate one-time improvements in performance by limiting the population of test takers, but will not necessarily increase its likelihood of success in future years. But the same may be true of many other worthwhile pedagogical reforms. (Consider what would happen if academics were rewarded based upon the increased number of articles published from one year to the next, rather than some average of the stock of accumulated work and the average output per year over their careers.)

Second, we know from the optimal-incentives literature summarized in Lazear (1995) and Milgrom and Roberts (1992) that imperfect measures of performance should receive less weight in an incentive framework. Test-score measures are imperfect measures of schools' output for at least four reasons.

First, test-score measures often include systematic, predictable factors that are outside schools' control. The easiest example of these factors is family background. Placing too great an implicit weight on family background and other factors affecting students' baseline performance encourages schools to exempt students from their testing programs. One partial solution to this problem is to focus on gain scores or value-added measures of achievement (it is only a partial solution since some students
not only start out with a lower baseline, but they may have a predictably flatter or steeper trajectory as well).

Second, as the authors note, the typical test-based measures are incomplete measures of school output. For example, most test-based accountability systems are based upon reading and math scores alone. As critics are wont to point out, civics and social tolerance are typically assigned zero value. However, it is also worthwhile to note that many "hard" skills—such as science, history, and social studies—are also excluded. The new federal No Child Left Behind Act of 2001 requires states to test reading and math skills in grades three through eight by the 2005–06 school year. Science tests will not be added until 2007–08. There are no plans to require states to test other skills.

Placing too great a weight on the measured outputs is likely to lead schools to substitute away from other valued, but difficult-to-measure domains. Whether intended or not, such rules are likely to tip the balance of instruction toward the subset of subject areas and concepts that are tested. For example, in the Kentucky accountability system in the early 1990s, science was tested in fourth grade and math was tested in fifth grade. Stecher and Barron (1999) found that teachers had reallocated their time so that they spent more time on science in fourth grade when students took the science test and more time on math in fifth grade when students took the math test. Jacob (2002) found that scores on science and social studies leveled off or declined in Chicago after the introduction of an accountability system that focused on math and reading performance.

Third, school-level test scores are also imprecise measures of the domains they are intended to measure. This fact is highlighted by Figure 1, which reports the distribution of different types of measures by school size, taken from a North Carolina sample. Panel A reports data on mean math performance in grades three through five by school size; Panel B reports data on changes in mean performance in grades three through five by school size; the final panel reports mean gains in performance at the individual student level in grades four and five. As is evident in the funnel-shaped patterns for all three distributions in Figure 1, one important source of imprecision is simple sampling variation. Given that the typical elementary school contains 60 students per grade level, a few particularly bright or particularly rowdy students can have a big impact on scores from year to year. Aggregating across several grades helps, but obviously does not eliminate this problem. Moreover, sampling variation appears to account for a larger share of the total variance for the change in performance from year to year and for the mean cohort gain across different schools than for levels.

Fourth, in addition to sampling variation, there is evidence of other one-time shocks to school performance (Kane and Staiger 2002a). These shocks may be due to other sampling-related causes—such as peer effects, testing artifacts generated by changes in test forms, school-wide distur-

#### Figure 1 Distribution of Levels, Changes, and Cohort Gains, by School Size



Panel B: Change in Mean Math Performance, Grades 3 to 5, by Total Number of Students in Grades 3 to 5



**Panel C:** Mean Gain in Math Performance, Grades 4 and 5, by Total Number of Students in Grades 4 and 5





bances such as a dog barking in the parking lot on the day of the test, or other short-term impacts such as classroom chemistry. The pattern of such shocks suggests that there is a weak correlation in performance between test scores one year apart, but that correlation fades only gradually after that one year. Such one-time shocks are unlikely to be due to teacher turnover, since teacher turnover follows a very different pattern. After one year, about 20 percent of teachers turned over in the typical elementary school in North Carolina. However, after five years, about 50 percent of the teachers in a school had turned over. Teacher turnover may explain the pattern of declining correlation in the second year and beyond, but it cannot explain the dramatic fall-off in the first year.

Kane and Staiger (2002a) performed further analysis of the sources of variance in test scores in North Carolina, including decomposing the variance in status measures ("levels"), status-change measures ("changes"), and cohort-gain measures ("gains") into three parts: a persistent component, sampling variation, and other one-time shocks. They report four results worth noting: First, the between-school variance in mean test performance is small relative to the total variance in performance at the student level. Even including the effect of sampling variation, the between-school variance accounted for only 10 percent to 20 percent of the total variance in test scores. Despite the fact that there may be some very high-scoring schools and some very low-scoring schools, the differences in performance for students within the typical school tend to be much larger than the differences between schools.

Second, much of the difference in the test-score levels is persistent. Even among the smallest quintile of schools, nonpersistent factors account for only 27 percent of the variance between schools. Among the largest quintile of schools, such factors account for only 13 percent of the variance. However, since we are not adjusting for initial performance levels or for the demographic characteristics of the students, much of that reliability may be due to the unchanging characteristics of the populations feeding those schools and not necessarily from unchanging differences in school performance.

Third, although one might be tempted to rate schools by their improvement in performance or by the average increase in student performance over the course of a grade, such attributes are measured remarkably unreliably. More than half (56 percent) of the variance among the smallest quintile of schools in mean gain scores is due to sampling variation and other nonpersistent factors. Even among the largest quintile of schools, nonpersistent factors are estimated to account for 34 percent of the variance in gain scores. Changes in mean test scores from one year to the next are measured even more unreliably. More than 80 percent of the variance in the annual change in mean test scores among the smallest quintile of schools is due to one-time, nonpersistent factors.



Fourth, increasing the sample size by combining information from more than one grade will do little to improve the reliability of changes in test scores over time. Even though the largest quintile of schools was roughly four times as large as the smallest quintile, the proportion of the variance in annual changes caused by nonpersistent factors was still over 60 percent.

Kane and Staiger (2002a) develop several implications of such imprecision for the design of accountability systems. Figure 2 illustrates the impact of imprecision in test-score measures on schools' incentives. Suppose a small school and a large school have the same expected performance next year. Each has a range of expected outcomes. Suppose that only those schools with scores above a threshold will win an award. The marginal incentive for each school is measured by the height of the density function where it crosses the incentive. For thresholds up at the extremes, more randomness can actually increase the strength of incentives. In this picture, when the threshold is at either extreme, small schools have a positive incentive to improve, while large schools have very little incentive. When the threshold is in the middle of the distribution, small schools with a greater variance in likely scores have the weaker incentive.

## HOW MUCH WOULD AN INCREASE IN PERFORMANCE BE WORTH?

Critics of school accountability worry that current systems already place too great a weight on imperfect measures of academic achievement and, on net, may do more harm than good. To evaluate these concerns, one must have a sense of the potential value that we should place on an increase in student achievement.

Some simple calculations by Kane and Staiger (2002b) reveal that the monetary value of even a small improvement in academic achievement can have very large payoffs. Two recent papers provide estimates of the impact of test performance on the hourly wages of young workers. Murnane, Willett, and Levy (1995) estimate that a one-standard-deviation difference in math test performance is associated with an 8.0 percent hourly wage increase for men and a 12.6 percent increase for women. These estimates probably understate the value of test performance, since the authors also control for years of schooling completed. Neal and Johnson (1996), who do not condition on educational attainment, estimate that a one-standard-deviation improvement in test performance is associated with hourly wage increases of 18.7 percent for men and 25.6 percent for women. Using a discount rate of 6 percent, the present value at age 18 of a one-standard-deviation difference in test performance is worth roughly \$62,000 per student using the Murnane, Willett, and Levy estimates and \$146,000 per student using the higher estimates from Neal and Johnson.<sup>1</sup> Discounting these values back to age 9 (for example, fourth grade) would reduce the estimates to \$40,000 and \$94,000 per student.

Such estimates are quite large relative to the rewards offered to schools for increasing student test performance. For example, California paid elementary schools and their teachers an average award of \$122 per student if their school improved student performance by an average of at least 0.03 student-level standard deviation.<sup>2</sup> Based on the calculations in

PV at Age 18 = 
$$\sum_{i=1}^{46} \beta w_i \left(\frac{1+\gamma}{1+r}\right)^{i-1}$$
,

where  $\beta$  is the proportional rise in wages associated with a given test-score increase; w<sub>i</sub> represents wages from age 18 through 64 estimated using full-time, year-round workers in the 2000 Current Population Survey;  $\gamma$  represents the general level of productivity growth, assumed to equal 0.01; and r is the discount rate, assumed to equal 0.06.

<sup>2</sup> The School Site Employee Bonus program provided \$591 per full-time equivalent teacher to both the school and teacher, or \$59 per student based on an average of 20 students per teacher. The Governor's Performance Award (GPA) program provided an additional \$63 per student. The growth target for the average elementary school was 9 points on the state's academic performance index (API). Because the state did not publish a student-level

<sup>&</sup>lt;sup>1</sup> I used the following calculation:

the preceding paragraph, the present value of such an increase in test scores to students in elementary school would be in the range of \$1,200 to \$2,800 per student (0.03 times \$40,000, or \$94,000), much more than the \$122 paid by the state. In other words, the labor-market value of the test-score increase would have been worth roughly 10 times to 20 times the value of the incentive provided in 2001 by California—the state with the most aggressive financial incentive strategy in that year. (Budget cuts have subsequently led to declines in those incentive payments.)

This calculation suggests that even the most aggressive state is paying schools much less than the marginal payoff if we thought the test-score improvements reflected true achievement. Critics' concerns about relying on imperfect performance measures may already be reflected in small incentive payments. In fact, the strength of incentives for schools in California is similar to what Hall and Liebman (1998) found for CEOs: \$1 in compensation for every \$40 increase in firm valuation.

## EMPIRICAL ESTIMATES OF THE IMPACT OF ACCOUNTABILITY INCENTIVES ON TEST PERFORMANCE

The most intriguing part of the Hanushek–Raymond paper studies the relationship between state differences in the timing of adoption of test-based accountability and state performance on the NAEP. States with an accountability system in place in 2000 had achievement growth approximately 1 percent larger than states without such systems. The number of years a state had such a system in place was not related to NAEP performance.

Hanushek and Raymond report the impact in log units, not in student-level standard deviation units. A few simple calculations suggest that the impact was fairly modest when translated into standard deviation units. Between 1996 and 2000, the average growth in achievement on the state assessments between fourth and eighth grade was 52 points (from 222 to 274). A 1 percent difference, therefore, would represent a 0.5 point increase. The standard deviation in achievement in fourth or eighth grade was approximately 32 points. Therefore, a 1 percent improvement in the growth in performance from fourth to eighth grade would

standard deviation in the API scores, we had to infer it. A school's API score was a weighted average of the proportion of students in each quintile of the national distribution on the reading, math, language, and spelling sections of the Stanford 9 test. For elementary schools, the average proportion of students across the four tests in each quintile (from lowest to highest) was 0.257, 0.204, 0.166, 0.179, and 0.194, and the scores given to each quintile were 200, 500, 700, 875, and 1,000. Under the assumption that students scored in same quintile or all four tests, we could calculate the student-level variance as 0.257 (200 - 620)<sup>2</sup> + 0.204 (500 - 620)<sup>2</sup> + 0.179 (875 - 620)<sup>2</sup> + 0.194 (1,000 - 620)<sup>2</sup> = 89,034, implying a standard deviation of 298. This is nearly five times the school-level variance, which is roughly consistent with expectations.

represent a 0.016 student-level standard deviation improvement in performance for the average student. However, given the estimates above, even a small increase in performance may well be worthwhile. For elementary school students, a 0.016 student-level standard deviation increase would be worth \$640 to \$1,500 per student. Most states are spending much less than that on their accountability systems. Therefore, a more thorough cost-benefit analysis may yield quite large payoffs to creating an accountability system.

As Hanushek and Raymond acknowledge, accountability systems are weakened if an increasing number of students are excluded from taking the exams. We should be cautious in using the NAEP tests to study the impact of state accountability systems because there have been large increases over time in the proportion of students excluded from the state NAEP samples.<sup>3</sup> The NAEP test has traditionally excluded the test scores of students to whom the states have granted testing accommodations such as allowing a longer time to take the test, having the questions read aloud, or having the test translated into a native language. The idea was to compare students in the same testing conditions. However, after the passage of the Individuals with Disabilities Act in 1996, many states began granting accommodations to a larger share of students. (There may be more nefarious reasons as well; now that NAEP scores are given a much higher profile, states have a stronger incentive to inflate their scores by excluding students.)

Moreover, the increases in exclusion rates seem to be particularly large in states that have been touted for their increases in NAEP performance. Figure 3 reports the change in eighth-grade math NAEP scores between 1992 and 2000 and the changes in the proportion of the sample excluded from the assessments. Between 1992 and 2000, the average state increased its exclusions by 3.5 percentage points, from 5 percent of sampled youth to 8.5 percent of sampled youth. One state, often cited as having an exemplary accountability system, North Carolina, increased its eighth-grade exclusion rate by 11 percentage points, more than in any other state.<sup>4</sup>

Because of this data problem, we may never be able to go back and assess the considerable experimentation with accountability initiatives that occurred in many states during the 1990s. Beginning with the 2000

<sup>&</sup>lt;sup>3</sup> Grissmer and Flanagan (2002) note the same phenomenon.

<sup>&</sup>lt;sup>4</sup> It is unlikely that the change in exclusion rates accounts for all of the change in North Carolina. The exclusion rates in fourth grade and in eighth grade together increased an average of 10 percentage points. If the distribution of test scores is normal at the student level, then raising the truncation point from the 3rd percentile to the 13th percentile would have raised test scores by only 0.17 standard deviation—much less than the observed increase in North Carolina. This is an extreme assumption since not all of the nontested students would have been in the bottom tail, so that the actual effect on NAEP scores is probably smaller.



assessment, the NAEP began reporting state-level results for the "no accommodations" sample as well as for a sample including the students with accommodations. In the future, then, it may be easier to track differences in improvements at the state level—although there will still be a tricky problem created by the fact that different states will continue to grant accommodations to different shares of their students.

## CONCLUSION

Hanushek and Raymond provide an extremely useful description of state accountability schemes and review the developing literature on the impact of test-based accountability on academic achievement. Their analysis of the growth in NAEP scores in states with and without accountability systems suggests small, positive impacts on student performance. It is worthwhile noting that even a small increase in student performance would generate sufficient benefits to cover the moderate cost of operating an accountability system, given the value of academic achievement to students later in life.

Given the range of strategies used in different states—some states reward test-score levels, while other states reward changes in test scores, while still other states focus on cohort-gain scores—it is clear that we have a lot to learn about the relative payoffs of different approaches. Therefore, it is unfortunate that the No Child Left Behind Act of 2001 imposes a new federal system of accountability that will inevitably conflict with many of the state rating systems in use. Under that system, states are allowed to define proficiency in whatever manner they choose. But once a state has defined proficiency, the minimum proficiency rate for all schools and for all racial and ethnic subgroups within schools will be equal to the proficiency rate of the 20th percentile school. Because the federal system will be based on status measures (or levels), while many states use status changes or gain scores, there will be many cases where schools are failing the federal definition while doing well using their state's metric. Many schools that fare well under California's system based on changes in test performance or under North Carolina's system using cohort-gain scores, even many of those achieving exemplary rankings, will be sanctioned under the new federal law. It remains to be seen whether the mixed signals created when the new federal accountability system is laid on top of state accountability systems will simply confuse schools and parents or whether it will spur them on to further improvements.

#### References

- Grissmer, David and Ann Flanagan. 2002. "Tracking the Improvement in State Achievement Using NAEP Data." RAND Corporation, unpublished paper.
- Hall, Brian J. and Jeffrey B. Liebman. 1998. "Are CEO's Really Paid Like Bureaucrats?" Quarterly Journal of Economics 113 (3): 653–91.
- Jacob, Brian A. 2002. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." NBER Working Paper No. 8968 (May).
- Kane, Thomas J. and Douglas O. Staiger. 2002a. "Volatility in School Test Scores: Implications for Test-Based Accountability Systems." In *Brookings Papers on Education Policy*, 2002, edited by D. Ravitch. Washington, DC: Brookings Institution.
- ——. 2002b. "The Promise and the Pitfalls of Using Imprecise School Accountability Measures." Journal of Economic Perspectives 16 (4): 91–114.
- Lazear, Edward P. 1995. Personnel Economics. Cambridge, MA: MIT Press.
- Milgrom, Paul and John Roberts. 1992. *Economics, Organization and Management*. Englewood Cliffs, NJ: Prentice Hall.
- Murnane, Richard J., John B. Willett, and Frank Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." *Review of Economics and Statistics* 77 (2): 251–66.
- Neal, Derek and William Johnson. 1996. "The Role of Premarket Factors in Black–White Wage Differentials." *Journal of Political Economy* 104 (5): 869–95.
- Stecher, Brian and Sheila Barron. 1999. "Quadrennial Milepost Accountability Testing in Kentucky." CSE Technical Report No. 505. Los Angeles: Center for the Study of Evaluation, Standards, and Testing.

# WHAT IS THE APPROPRIATE ROLE FOR STUDENT ACHIEVEMENT STANDARDS?

## John H. Bishop\*

Three presidents, the National Governors Association, and numerous blue-ribbon panels have called for the development of state content standards for core subjects and examinations that assess the achievement of these standards. The Competitiveness Policy Council, for example, advocates that "external assessments be given to individual students at the secondary level and that the results should be a major but not exclusive factor qualifying for college and better jobs at better wages" (1993, p. 30). The American Federation of Teachers advocates a system in which "students are periodically tested on whether they're reaching the standards, and if they are not, the system responds with appropriate assistance and intervention. Until they meet the standards, they won't be able to graduate from high school or enter college" (American Federation of Teachers 1995, pp. 1–2).

American policymakers are trying to deal with low standards and weak incentives for hard study by making students, staff, and schools more accountable for learning. The education departments of the 50 states have responded by developing content standards for core academic subjects, administering tests assessing this content to all students, pub-

<sup>\*</sup> Associate Professor of Human Resource Studies, New York State School of Industrial and Labor Relations, Cornell University. The preparation of this paper was made possible by support from the Center for Advanced Human Resource Studies and the Consortium for Policy Research in Education (funded by the Office of Educational Research and Improvement, U.S. Department of Education). The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education. This paper has not been formally reviewed or approved by the faculty of the ILR School. It is intended to make results of Center research, conferences, and projects available to others interested in human resource management in preliminary form to encourage discussion and suggestions.

lishing individual school results, and holding students and schools accountable for student achievement. While these efforts are generically referred to as standards-based reform, the mix of initiatives varies a great deal from state to state.

It is claimed that a curriculum-based external exit exam (CBEEE) system based on world-class content standards will improve the teaching and learning of core subjects. What evidence is there for this claim? What impacts have such systems had on school policies, teaching, and student learning? Outside the United States, CBEEE systems are the rule, not the exception. Within the United States, New York's Regents exams and North Carolina's end-of-course (EOC) exams are two examples of such systems. Do New York and North Carolina students outperform students with similar socioeconomic backgrounds from other states?

## CURRICULUM-BASED EXTERNAL EXIT EXAMINATION SYSTEMS

While a number of states-for example, Maryland, Mississippi, Oklahoma, Arkansas, Tennessee, Virginia, Michigan-appear to be planning to implement CBEEE systems, only two states-New York and North Carolina—had established such systems by the beginning of the 1990s. State-sponsored end-of-course exam systems are provided in Appendix Table 1. The granddaddy of these is New York's Regents exam system. It has been in continuous operation since the 1860s. Panels of local teachers grade the exams using rubrics supplied by the state Board of Regents. Exam scores appear on transcripts and are the final exam mark that is averaged with the teacher's quarterly grades to calculate the final course grade. A college-bound student taking a full schedule of Regents courses would typically take Regents exams in mathematics and earth science at the end of ninth grade; mathematics, biology, and global studies at the end of tenth grade; mathematics, chemistry, American history, English, and foreign language at the end of eleventh grade; and physics at the end of twelfth grade. However, taking Regents courses and, therefore, Regents exams was voluntary until late in the 1990s. Prior to 1998, nearly half of the students chose to take "local" courses originally intended for noncollege-bound students, knowing that good grades could be obtained without much effort.

Between 1987 and 1991, North Carolina introduced end-of-course exams for Algebra 1 and 2, Geometry, Biology, Chemistry, Physics, American History, Social Studies, and English 1. Versions of these courses that are not assessed by a state test do not exist, so virtually all North Carolina high school students take at least six of these exams. Test scores appear on the student's transcript, and most teachers have been incorporating EOC exam scores in course grades. Starting in the year 2000, state law requires the EOC tests to have at least a 25 percent weight in the final course grade. Clearly from this description one can see that even North Carolina's EOC exams and New York's Regents exams prior to 1999 carried only low to moderate stakes for students.

## **MINIMUM COMPETENCY EXAMINATIONS**

Most states pursuing standards-based reform have established minimum competency exam (MCE) or other test-based school accountability systems that are quite different from curriculum-based external exit exam systems. Appendix Table 2 presents information on the end-of-grade (EOG) examination systems that a number of states have adopted, often to determine eligibility for honors diplomas or scholarships.

Eighteen states have MCE graduation requirements, and another 11 states are developing or phasing in MCEs. Minimum competency exams raise standards, but probably not for everyone.<sup>1</sup> The standards set by the teachers of honors classes and advanced college prep classes are not changed by an MCE. Students in these classes generally pass the MCE on the first try without special preparation. The students who are in the school's least-challenging courses experience the higher standards. Students pursuing a "do the minimum" strategy are told "you must work harder" if you are to get a diploma and go to college. School administrators want to avoid high failure rates, so they are likely to focus additional energy and resources on raising standards in the early grades and improving the instruction received by struggling students.

## SCHOOL REPORT CARDS AND STAKES FOR TEACHERS AND ADMINISTRATORS

Formal systems for holding schools accountable are growing in popularity. In 1999, 37 states were publishing school report cards for all or almost all of their schools (Edwards 1999). Publicly identifying low-performing schools is intended to spur local school administrators and boards of education to undertake remedial action. Nineteen states had a formal mechanism for rewarding schools either for year-to-year gains in achievement test scores or for exceeding student achievement

<sup>&</sup>lt;sup>1</sup> Minimum competency exams are in addition to—not a replacement for—teacherimposed standards. In an MCE regime, teachers continue to control the standards and assign grades in their own courses. Students must still get passing grades from their teachers to graduate. The MCE regime imposes an additional graduation requirement and thus cannot lower standards (Costrell 1994). The Graduate Equivalency Diploma (GED), by contrast, offers students the opportunity to shop around for an easier (for them) way to a high school graduation certificate. As a result, the GED option lowers overall standards. This is reflected in the lower wages that GED recipients command (Cameron and Heckman 1991).

targets (Edwards 1999). Nineteen states had special assistance programs to help failing schools turn themselves around. If improvements were not forthcoming, 11 states had the power to close down, take over, or reconstitute failing schools.

Exactly how are these student and school accountability systems similar to or different from the curriculum-based external exit exam systems that are found abroad and in New York and North Carolina? We begin by noting the features they have in common. The following five criteria apply to CBEEEs and MCEs:

1. The exams produce signals of accomplishment that have real consequences for students and schools. While some stakes are essential, high stakes may not be necessary. Analyses of Canadian and U.S. data summarized below suggest that moderate stakes may be sufficient to produce substantial increases in learning.

2. The exams define achievement relative to an external standard, not relative to other students in the classroom or the school. Fair comparisons of achievement across schools and across students at different schools are possible. Costrell's (1994) analysis of the optimal setting of educational standards concluded that more centralized standard-setting (state or national achievement exams) results in higher standards, higher achievement, and higher social welfare than decentralized standard-setting (in other words, teacher grading or schools' graduation requirements).

**3.** The exams assess a major portion of what students are expected to know and be able to do. Studying to prepare for an exam (whether set by one's own teacher or by a state department of education) should result in the student's learning important material and developing valued skills. Some MCEs, CBEEEs, and teacher exams do a better job of achieving this goal than others. External exams, however, cannot assess every instructional objective. Teachers themselves must accept responsibility for evaluating dimensions of performance that cannot be reliably assessed by external means or that local leaders want to add to the learning objectives specified by the state department of education.

4. The exams cover all or almost all students. Exams for elite schools, advanced courses, or college applicants will influence standards at the top of the vertical curriculum, but will probably have limited effects on the rest of the students. With MCEs, in contrast, virtually all students are affected, and the school system as a whole must accept responsibility for how students do on the exams. A single exam taken by all is not essential. Many nations allow students to choose which subjects to be examined in and offer high- and intermediate-level exams in the same subject.

5. The exams are controlled by the education authority that establishes the curriculum for and funds K–12 education. Curriculum reform is facilitated because coordinated changes in instruction and

exams are feasible. Tests established and mandated by other organizations serve the interests of other masters. America's premier high-stakes exams—the SAT-I and the ACT—serve the needs of colleges to sort students by aptitude, *not* the needs of schools to reward students who have learned what high schools are trying to teach.

Curriculum-based external exit exam systems are distinguished from MCEs by the following additional features:

**6.** The system signals multiple levels of achievement in the subject. If only a pass/fail signal is generated by an exam, and passing is necessary to graduate, the standard will almost inevitably be set low enough to allow almost everyone to pass after multiple tries. The great bulk of students will easily pass the test and will have no incentive to strive to do better. CBEEEs, in contrast, signal the student's achievement level in the subject being tested, so that all students, not just those at the bottom of the class, have an incentive to study hard in order to do well on the exam. Consequently, a CBEEE should be more likely to improve classroom culture than an MCE. Costrell agrees: "The case for perfect information [making scores on external examinations available rather than just whether the individual passed or failed] would appear to be strong, if not airtight: for most plausible degrees of heterogeneity, egalitarianism, and pooling under decentralization, perfect information not only raises GDP, but also social welfare" (1994, p. 970).

**7. The system assesses more difficult material.** Since CBEEEs are supposed to measure and signal the full range of achievement in a subject, they contain more difficult questions and problems. This induces teachers to spend more time on cognitively demanding skills and topics. MCEs, by contrast, are designed to identify which students have failed to surpass a rather low minimum standard, so they do not ask questions or set problems that students near that borderline are unlikely to be able to answer or solve.<sup>2</sup> This tends to result in excessive class time being devoted to practicing low-level skills.

8. The system is a collection of end-of-course exams. Since CBEEEs assess the content of a specific course, teachers of the course (or course sequence) being tested inevitably will feel responsible for how well their students do on the exam. Grades on EOC exams may be made part of the overall course grade, further integrating the external exam into the

<sup>&</sup>lt;sup>2</sup> In 1996, only 4 of the 17 states with MCEs targeted their graduation exams at a tenth-grade proficiency level or higher. Failure rates for students taking the test for the first time varied a great deal: from highs of 46 percent in Texas, 34 percent in Virginia, 30 percent in Tennessee, and 27 percent in New Jersey to a low of 7 percent for Mississippi. However, since students can take the tests multiple times, eventual pass rates for the class of 1995 were much higher: 98 percent in Louisiana, Maryland, New York, North Carolina, and Ohio; 96 percent in New Jersey; 91 percent in Texas; and 83 percent in Georgia (American Federation of Teachers 1996).

classroom culture. Alignment between instruction and assessment is maximized, and accountability is enhanced. Proponents argue that teachers will not only want to set higher standards, but will also find their students more attentive in class and more likely to complete demanding homework assignments. Teachers become coaches helping their team battle the state exam.

Those who are skeptical about the value of introducing CBEEEs point out that American students already take a lot of standardized tests. Why aren't the tests students already take (such as the ACT, the SAT-I, or commercially prepared norm-referenced achievement tests) sufficient? What's so special about the new CBEEEs that some states are introducing in their standards-based reforms?

Norm-referenced achievement tests such as the CAT, CTBS, ITBS, ITED, and Terra Nova are not curriculum-based external exit exams because they fail criteria one and eight. Students have no stake in doing well on these tests. They are not part of a course grade or important to the student in some other way, so many high school students fail to put much effort into answering all the questions correctly and completely.<sup>3</sup> Where stakes are not attached to results, teachers and school administrators experience the consequences, rather than individual students. In most of the nation, tests that students have no reason to try hard on are the primary indicator of student achievement in school accountability systems. When this is the case, school ratings may reflect the school's success in getting students to try hard on state tests and not the quality of instruction throughout the school year. This reduces the validity of high school accountability systems problematic.

The SAT-I test is not a CBEEE because it does not fulfill criteria three, five, and eight. It fails to assess most of the material—history, science, economics, civics, literature, foreign languages, and the ability to write an essay—that high school students are expected to learn. The Scholastic Aptitude Test (SAT) was designed from the beginning to minimize

<sup>&</sup>lt;sup>3</sup> This observation is based on interviews with the directors of the testing and accountability divisions in Manitoba and New Brunswick. It is also based on the large increases in student performance that occurred in New Brunswick, Massachusetts, Michigan, and other states when no-stakes tests became moderate- or high-stakes tests (Hayward 2001). Experimental studies confirm the observation. In Candace Brooks-Cooper's master's thesis (1993), a test containing complex and cognitively demanding items from the NAEP history and literature tests and the adult literacy test was given to high school students recruited to stay after school by the promise of a \$10.00 payment for taking the test. Students were randomly assigned to rooms. Students in one room were promised a payment of \$1.00 for every correct answer greater than 65 percent correct. This group did significantly better than the other students, who were told different test-taking conditions, including the standard "try your best" condition. Similar results were obtained in other well-designed studies conducted by the National Center for Research on Evaluation, Standards and Student Testing (see Kiplinger and Linn 1993 and O'Neil et al. 1997).

backwash effects on teaching and student study habits. Indeed, when the machine-scored, multiple-choice SAT replaced the curriculum-based essay-style College Board Examinations, Harvard College's admissions director Richard Gummere was very candid about why the SAT had been adopted: "Learning in itself has ceased to be the main factor [in college admissions]. The aptitude of the pupil is now the leading consideration" (Gummere 1943, p. 5).

The subject-specific SAT-II achievement tests fail criteria one, four, and five. Stakes are very low—few colleges consider SAT-II results in admissions decisions—and few students take them. In 1982–83, only 6 percent of SAT-I test takers took a science SAT-II, and only 3 to 4 percent took one in history or a foreign language. Schools do not assume responsibility for preparing students for SAT-II tests.

The Advanced Placement (AP) examinations are the one exception to the generalization that the United States lacks a national CBEEE. The number of students taking AP examinations has been growing at a compound annual rate of 9 percent per year. In 1999, 686,000 students, about 11 percent of the nation's juniors and seniors, took at least one AP exam. Despite this success, however, 44 percent of high schools do not offer even one AP course, and many that do allow only a tiny minority of their students to take these courses (College Board 1999). Low participation means that AP exams fail criterion 5 and, consequently, are not a CBEEE system. They can, however, serve as a component of a larger system.

## HOW ARE CBEEE SYSTEMS HYPOTHESIZED TO INCREASE ACHIEVEMENT?

Curriculum-based external exit exam systems fundamentally change the signaling of student achievement. In doing so, they transform the incentives faced by students, parents, teachers, and school administrators. CBEEE systems are, consequently, hypothesized to influence the resources made available to schools and the priorities of school administrators, teacher pedagogy, parental encouragement, and student effort.

#### Impact on Students

Curriculum-based external exit exam systems improve the signaling of academic achievement. As a result, colleges and employers are likely to give greater weight to academic achievement when they make admission and hiring decisions, so the rewards for learning should grow and become more visible. CBEEE systems also shift attention toward measures of absolute achievement and away from measures of relative achievement, such as rank in class and teacher grades. In doing so, CBEEE systems ameliorate the problem of peer pressure against studying.

How serious a problem is peer pressure against studying? Steinberg, Brown, and Dornbusch's 1996 study of nine high schools in California and Wisconsin suggests that academic excellence is still not highly valued by peers in most schools:

The adolescent peer culture in America demeans academic success and scorns students who try to do well in school . . . less than 5 percent of all students are members of a high-achieving crowd that defines itself mainly on the basis of academic excellence. . . . Of all the crowds, the "brains" were the least happy with who they are—nearly half wished they were in a different crowd (pp. 16, 145–6).

Why do so many "brains" want to get out of their crowd? Don Merten's 1996 ethnography of Cronkite Junior High School provides a rich and perceptive description of why this is so. Documenting the thoughts and actions of the ostracized and the popular students, he describes the transformation of one student from outcast to socially acceptable classmate. His description of the student's journey from nerd to cool kid is a gripping illustration of the power of peer norms in middle school. In order to fit in, the student cast away the norms and values he had lived by in elementary school and had defended in seventh grade: empathy, helping others, being good. He adopted instead the more predatory anti-teacher persona promoted by the dominant/popular students in junior high school.

Unfortunately, the peer pressure against studying or excelling in school found in Cronkite Junior High School is not an aberration. In the Educational Excellence Alliance survey, 24 percent of students said, "My friends make fun of people who try to do real well in school." Fifty-six percent said, "My friends joke around and annoy the teacher."

The teachers and principals of many American middle schools have lost normative hegemony. In the eyes of most students, the "brains" exemplify the "I trust my teacher to help me learn" attitude that prevails in most elementary school classrooms. The dominant middle school crowd is saying that trusting teachers is baby stuff. It's "us" versus "them." Withdraw from alliances with teachers, they say, and get with the program of becoming popular with peers. Be like us, the popular crowds say. Spend your time socializing. Do not study too hard. Value classmates for their athletic prowess and their attractiveness, not their interest in history or their accomplishments in science.

Why are studious students treated as outcasts? In part, it is because exams are graded on a curve. When exams are graded on a curve or college admissions are based on rank in class, joint welfare is maximized if no one puts in extra effort. In the game that results, side payments (friendship and respect) and punishments (ridicule, harassment, and ostracism) enforce the cooperative "don't study" solution.

If, by contrast, students are evaluated relative to an outside standard, as they would be with CBEEEs, they no longer have a personal interest in getting teachers off track or persuading each other to refrain from studying. There is less incentive for them to engage in peer pressure that demeans studiousness.

#### Impact on School Administrators

When there is no external assessment of academic achievement, students and their parents benefit little from administrative decisions that opt for higher standards, more-qualified teachers, or a heavier student workload. The immediate consequences of such decisions—higher taxes, more homework, having to repeat courses, lower grade point averages (GPAs), complaining parents, a greater risk of being denied a diploma—are all negative.

When student learning is not assessed externally, the positive effects of choosing academic rigor are negligible and postponed. If college admission decisions are based on class rank, GPA, and aptitude tests not externally assessed achievement in secondary school courses upgraded standards will not improve the college admission prospects of next year's graduates. Graduates will probably do better in difficult college courses and will be more likely to get a degree, but that benefit is uncertain and far in the future. Maybe over time, the school's reputation and, with it, the college admission prospects of graduates will improve because the current graduates are more successful in local colleges. That, however, is even more uncertain and postponed. Publishing data on the proportions of students meeting targets on standardized tests probably speeds the process by which real improvements in a school's performance influence its local reputation. However, other indicators (such as SAT test scores, proportions going to various types of colleges, and the socioeconomic background of the students) tend to be more prominent. As a result, school reputations are determined largely by things over which teachers and administrators have little control.

American employers historically have paid little attention to student achievement in high school or school reputations when selecting young workers (Bishop 1990, 1992 and Hollenbeck and Smith 1984). Those that do pay attention to achievement use indicators of relative performance such as GPA and class rank rather than results on an external exam as a hiring criterion. Consequently, higher standards do not benefit students as a group, so parents as a group have little incentive to lobby strongly for higher teacher salaries, higher standards, and higher school taxes.

higher teacher salaries, higher standards, and higher school taxes. External exams transform the signaling environment. Hiring better teachers and improving the school's science laboratories now yield a visible payoff—more students passing the external exams and being admitted to top colleges. This in turn is likely to lead to more spending on schools, more rigorous hiring standards for secondary school teachers, and a higher priority assigned to student learning in the allocation of school budgets.

Additionally, reform-minded administrators can use CBEEE results to inspire teachers to raise standards for all students. The superintendent of a suburban New York district that has been nationally recognized for raising student achievement levels observes: "[External validators like Regents exams] were the best and only way in which we could get teachers and staff to see themselves as others might see them and not just keep looking in the mirror and seeing themselves as they would like to see themselves" (author's interview with a superintendent of an all-Regents high school, August 1997).

#### Impact on Teachers

Curriculum-based external exit exams often have profound effects on teacher-student relationships and on the nature of the student peer culture. Teachers who have taught in environments with and without CBEEEs, as I have, sense the difference. When a proposal was put forward in Ireland to drop the nation's system of external assessments and have teachers assess students for certification purposes, the union representing Ireland's secondary school teachers reacted as follows:

Major strengths of the Irish educational system have been:

- (i) the pastoral contribution of teachers in relation to their pupils, and
- (ii) the perception of the teacher by the pupil as an advocate in terms of nationally certified examinations rather than as a judge.

The introduction of school-based assessment by the pupil's own teacher for certification purposes would undermine those two roles, to the detriment of all concerned.... The role of the teacher as judge rather than advocate may lead to legal accountability in terms of marks awarded for certification purposes. This would automatically result in a distancing between the teacher, the pupil, and the parent. It also opens the door to possible distortion of the results in response either to parental pressure or to pressure emanating from competition among local schools for pupils (Association of Secondary Teachers of Ireland 1990, p. 1).

Note how the Irish teachers feared that switching entirely to internal assessment would result in their being pressured to lower standards. For American teachers, such pressure is a daily reality. Thirty percent of American teachers say they "feel pressure to give higher grades than students' work deserves," and they "feel pressure to reduce the difficulty and amount of work you assign" (Peter D. Hart Research Associates 1995, p. 9). Under a system of external exams, teachers and local school

administrators lose the option of lowering standards to reduce failure rates and raise self-esteem. The only response open to them is to demand more of their students so as to maximize their chances of being successful on the external exams.

A further benefit of CBEEEs is the professional development that teachers receive when they are brought to centralized locations to grade the extended answer portions of examinations. In May 1996, I interviewed a number of teachers and union activists about the examination system in Alberta. Even though the union and these teachers opposed the exams, they universally shared the sentiment that serving on grading committees was "a wonderful professional development activity."<sup>4</sup> Having to agree on what constituted excellent, good, poor, and failing responses to essay questions or open-ended math problems resulted in a sharing of perspectives and teaching tips that most found very helpful. On the other hand, many fear that external exams will have a

On the other hand, many fear that external exams will have a negative effect on teaching. Opponents argue that "preparation for high-stakes tests often emphasizes rote memorization and cramming of students and drill and practice teaching methods" and that "some kinds of teaching to the test permit students to do well in examinations without recourse to higher levels of cognitive activity" (Madeus 1991, pp. 7–8). CBEEE advocates counter by challenging the assumption implicit in the above argument that examinations developed by the committees of

CBEEE advocates counter by challenging the assumption implicit in the above argument that examinations developed by the committees of teachers working for state departments of education are or will be worse than the tests developed by individual teachers. In fact, the tests that teachers develop for themselves are generally of very low quality. As John Thomas discussed at a 1991 conference, Fleming and Chambers's 1983 study of tests developed by high school teachers found that "80 percent of the items on teachers' tests were constructed to tap the lowest of [Bloom's] taxonomic categories: knowledge (of terms, facts, or principles)" (Thomas, p. 14). Rowher and Thomas (1987) found that only 18 percent of history test items developed by junior high teachers and 14 percent of items developed by senior high teachers required the integration of ideas. College instructors, by contrast, required such integration in 99 percent of their test items. Secondary school teachers test low-level competencies because that is what they teach.

If care is taken in designing external exams, they can induce improvements in instructional practice. Sherman Tinkelman describes one such instance, based on his experience as New York state's assistant commissioner for examinations and scholarships:

<sup>&</sup>lt;sup>4</sup> Interview results are available from the author upon request.

For years our foreign language specialists went up and down the state beating the drums for curriculum reform in modern language teaching, for change in emphasis from formal grammar to conversation skills and reading skills. There was not very great impact until we introduced, after notice and with numerous sample exercises, oral comprehension and reading comprehension into our Regents examinations. Promptly thereafter, most schools adopted the new curricular objectives (1966, p. 12).

## DO CBEEES INCREASE ACHIEVEMENT? A LOOK AT THE EVIDENCE

The hypothesis that curriculum-based external exit examination systems improve achievement can be tested by comparing nations, states, and provinces that do and do not have such systems. Here we examine five different data sets:

- science and mathematics achievement of eighth graders in 1995 and 1999 in the 50-nation Third International Math and Science Study (TIMSS);
- achievement of 14-year-olds in the Reading Literacy Study of the International Association for the Evaluation of Educational Achievement (IEA);
- science, mathematics, and reading literacy of 15-year-olds in the 2000 Program for International Student Assessment (PISA);
- science and mathematics scores of 13-year-olds in the International Assessment of Educational Progress (IAEP) for nine Canadian provinces; and
- SAT test results for New York state students compared with results for students in the rest of the United States.

The theory predicts that CBEEE systems influence administrators' decisions about school priorities, teachers' decisions about standards and pedagogy, and students' decisions about studying. Much of the ultimate impact of CBEEE systems on student achievement derives from the changes these systems induce in hiring decisions, school priorities, and teacher pedagogy. Bishop (1996) tested the effects of CBEEEs on most of these components using data on Canadian schools and students. In most of the analyses in the current paper, the units of observation are educational systems and the objective is to assess the total effect of CBEEE systems on student achievement. Total effects are estimated by a reduced-form model that controls for parental socioeconomic status, productivity, and national culture, not the endogenous administrator, teacher, and parent behaviors.

#### Third International Mathematics and Science Study (TIMSS)

TIMSS provides 1995 data for seventh and eighth graders for 40 countries. The TIMSS-Repeat study of eighth-grade achievement provides 1999 data for an additional 10 countries and a second measure of eighth-grade achievement for 25 countries. To determine which nations have curriculum-based external exit exams in secondary school, we reviewed comparative education studies, government documents, and education encyclopedias, and we interviewed education ministry officials, embassy personnel, and graduate students from those nations who were studying at Cornell University.<sup>5</sup>

The national school systems classified as having CBEEEs for both math and science in all parts of the country were Bulgaria, Czech Republic, Denmark, England, Finland, Hong Kong, Hungary, Indonesia, Iran, Ireland, Israel, Italy, Japan, Jordan, Korea, Lithuania, Malaysia, Morocco, the Netherlands, New Zealand, Poland, Russia, Scotland, Singapore, Slovak Republic, Slovenia, Taiwan, Thailand, Trinidad and Tobago, Tunisia, and Turkey. Three countries-France, Iceland, and Romania—had CBEEEs in mathematics but not in science. Four countries—Australia, Canada, Germany, and the United States—had CBEEEs in some states or provinces but not in others. Norway had regular exit examinations in mathematics, but exams only every few years in science. Latvia had an external examination system until very recently, so we gave it a 0.5 on the CBEEE variable. Sweden's unusual system of combining external assessment and teacher assessment was also assigned a 0.5. The countries classified as not having a CBEEE in either subject were Austria, Belgium (both Flemish- and French-speaking systems), Brazil, Chile, Colombia, Cyprus, Greece, Mexico, the Philippines, Portugal, Spain, Switzerland, and Venezuela.6

Countries with a CBEEE system in the subject tend to have higher TIMSS scores. Furthermore, achievement differentials across nations are

<sup>&</sup>lt;sup>5</sup> A bibliography of the documents and individuals consulted when making these classifications is available from the author upon request. The TIMSS report's information about examination systems does not distinguish between university admissions exams and curriculum-based exit exams, so its classifications are not useful for this exercise. The Philippines, for example, is classified as having external exams by the TIMSS report, but its exams are university admissions exams similar to the SAT. South Africa was excluded because its education system was disrupted for many years by anti-apartheid boycotts. Kuwait was excluded because of the disruption of its education system by the Iraqi invasion and the Gulf War.

<sup>&</sup>lt;sup>6</sup> Following Madeus and Kellaghan (1991), the university entrance examinations in Greece, Portugal, Spain, and Cyprus, and the ACT and SAT in the United States were not considered to be CBEEEs. University entrance exams have much smaller incentive effects because students who are headed into work do not take them, and teachers can avoid responsibility for their students' exam results by arguing that not everyone is college material or that examiners have set an unreasonably high standard to limit enrollment in higher education.

= = ; = = = = = = = = =	- ,	5 5			
	Curriculum- Based External Exit Exam	Log GDP per capita, 1999	East Asia	Adjusted R <sup>2</sup> RMSE	Number of Observations
TIMSS Science					
(U.S. GLE=26)					
8 <sup>™</sup> grade, 1995	51.1***	35.2***	6.2	.487	40
	(11.5)	(8.6)	(14.8)	32.1	
8 <sup>th</sup> grade, 1995–99	36.9***	57.8***	17.4	.521	50
	(12.1)	(8.0)	(14.1)	36.6	
TIMSS Math (U.S. GLE=24)					
8 <sup>th</sup> grade, 1995	42.3***	35.2***	54.2***	.484	40
0	(13.3)	(8.6)	(16.4)	36.3	
8 <sup>th</sup> grade, 1995–99	35.1***	65.3***	49.8***	.586	50
<b>J J J J J J J J J J</b>	(13.0)	(8.3)	(14.4)	37.9	
IEA Reading (U.S. GLE=24)	( )	()	( )		
Age-Adjusted	26.5***	29.1***	-17.6*	.610	25
Average, 14- year-olds, 1990	(8.1)	(9.0)	(11.8)	16.7	

#### Table 1

Academic Achievement in Nations with and without Curriculum-Based External Exit Examination Systems

TIMSS, TIMSS-Repeat, and IEA Reading Study Data

Note: Numbers in parentheses are *t*-statistics. TIMSS is the Third International Math and Science Study. GLE is grade-level equivalent. IEA is the International Association for the Evaluation of Educational Achievement. On a two-tail test, \*\*\* indicates p < 0.01, \*\* indicates p < 0.05, and \* indicates p < 0.10.

Source: Author's calculations using TIMSS, TIMSS-Repeat, and IEA Reading data. When test-score data were available for both 1995 and 1999, the dependent variable was the average of the two estimates. Gross domestic product per capita data are from World Bank (2001).

very large. According to the 1995 scores in science, Singapore, Korea, Bulgaria, and Flemish Belgium are more than one U.S. grade-level equivalent (GLE) ahead of the United States.<sup>7</sup> Colombia, the Philippines, Lithuania, Romania, and Portugal are more than three GLEs behind. In mathematics, Singapore, Korea, Japan, and Hong Kong are four or more GLEs ahead of the United States, while Colombia, the Philippines, and Iran are more than three GLEs behind.

We regressed the mean eighth-grade science and mathematics test scores on 1999 per capita gross domestic product deflated by a purchasing power parity price index, a dummy for East Asian nations, and a dummy for CBEEEs. The results of the analysis of TIMSS-95 scores are presented in the first and third rows of Table 1. The results of the analysis

<sup>&</sup>lt;sup>7</sup> A grade-level equivalent is defined here as the difference between seventh- and eighth-grade TIMSS test-score means for U.S. students. Overall, in the 1995 TIMSS, U.S. students ranked 15<sup>th</sup> in science and 31<sup>st</sup> in mathematics.

of merged TIMSS-95 and TIMSS-Repeat data are presented in the second and fourth rows. Both analyses indicate that test scores are significantly higher in more developed nations, East Asian nations, and nations with a CBEEE in the subject. Nations with CBEEEs are about 1.5 U.S. GLEs higher on the math and science tests in the combined TIMSS-95 and TIMSS-Repeat data. The differential is even larger when only the TIMSS-95 data are analyzed. Since exams are also likely to influence learning during upper secondary school, the total effect at the end of twelfth grade is likely to be larger still.

### International Association for the Evaluation of Educational Achievement (IEA) Reading Literacy Study

The IEA conducted a study of the reading literacy of 14-year-olds in 1990–91. The bottom row of Table 1 presents an analysis of IEA reading achievement data identical to the TIMSS analysis. The IEA study defined and measured three different types of reading literacy—narrative, expository, and document—and an average of the three scores is the dependent variable. The specification is the same as that used to study science and math achievement. The exam variable is an average of the TIMSS data. The IEA results are similar to the TIMSS results. Fourteen-year-old students in nations with CBEEE systems are about one U.S. grade-level equivalent better at reading than students in nations without CBEEE systems.

#### Program for International Student Assessment (PISA)

PISA is a new system of international assessment focusing on the reading, mathematics, and science literacy of 15-year-olds. Each participating country selected a nationally representative sample of approximately 4,000 15-year-olds. The students completed a 20- to 30-minute background questionnaire and a 90-minute assessment consisting of a mix of multiple choice, short answer, and extended response questions. PISA is a distinctive assessment tool: "While other studies, such as TIMSS and NAEP, have a strong link to curriculum frameworks and seek to measure students' mastery of specific knowledge, skills, and concepts, PISA is designed to measure 'literacy' more broadly. PISA's content is drawn from broad content areas, such as space and shape for mathematics, in contrast to more specific curriculum-based content such as geometry or algebra" (U.S. Department of Education 2001, p. 5).

Principals of schools where students took PISA assessments also completed a background questionnaire about their schools. PISA assesses the cumulative educational experiences of all students at age 15 regardless of their grade level or the type of institution they are attending. "By

0					
	Curriculum- Based External Exit Exam	Log GDP per capita, 1999	East Asia	Adjusted R <sup>2</sup> RMSE	Number of Observations
PISA 2000, 15-year-olds					
Science	30.6*** (9.9)	46.5*** (9.1)	43.2** (17.3)	.630 22.9	29
Math	38.3*** (12.7)	62.5*** (11.6)	40.5* (22.1)	.620 29.3	29
Combined Reading Literacy	31.8***	51.8*** (6.7)	10.3 (12.8)	.737 16.9	29
Retrieving Information	42.4*** (9.5)	64.1*** (8.2)	12.9 (15.8)	.747 20.7	29
Expected Years of Schooling, Ages 5–65					
Sum of Net Enrollment Rates	.08 (.61)	3.34*** (4.8)	.75 (.91)	.611 1.58	37
Sum of FTE Net Enrollment Rates	-0.8 (.46)	2.98*** (.36)	.47 (.69)	.698 1.20	37

#### Table 2

Academic Achievement in Nations with and without Curriculum-Based External Exit Examination Systems

Program for International Student Assessment 2000 Data

Note: FTE is full-time equivalent. Numbers in parentheses are *t*-statistics. On a two-tail test, \*\*\* indicates p < 0.01, \*\* indicates p < 0.05, and \* indicates p < 0.10.

Source: Expected years of schooling data are from Organization for Economic Cooperation and Development 2001 (p. 133); part-time enrollment counts as 0.5 year in full-time equivalent figures. PISA data are from U.S. Department of Education (2001).

assessing students near the end of compulsory schooling in key knowledge and skills, PISA provides information about how well prepared students will be for their future lives as they approach an important transition point for education and work" (U.S. Department of Education 2001, p. 3).

The first four rows of Table 2 present an analysis of PISA data on science, mathematics, and reading literacy. As in the TIMSS analysis, scores are significantly higher in more developed nations, East Asian nations, and nations with a CBEEE in the subject. While grade-level equivalents cannot be calculated for the PISA tests, estimated impacts of CBEEEs appear to be comparable to those in TIMSS and IEA reading studies. The effect of a CBEEE system is similar in magnitude to a doubling of a nation's productivity and income per capita.

These results are consistent with the causal hypotheses presented above. Causation is not proved, however, because other explanations can no doubt be proposed. Other sources of variation in curriculum-based exams need to be analyzed. Best of all would be studies that hold national culture constant; our final two data sets allow us to do this: the IAEP data for nine Canadian provinces, and SAT comparisons for New York state versus the other states.

Before turning to these last data sets, however, we can use the OECD data to see whether there is evidence that curriculum-based external exit exams tend to push students out of school. Many believe that a tradeoff exists between the standards and quality of an educational system and the number of students who can or will stay in school into their late teens and twenties. In the policy debate within the United States, concern has been expressed that high- or medium-stakes student accountability will increase dropout rates and reduce college attendance rates. We tested this hypothesis by calculating how many years youth in each of the OECD nations spend in school (we summed the net enrollment rates of people aged 5 to 65) and then assessing what impact CBEEEs have on these estimates of expected years of schooling. The results are presented in the fifth and sixth rows of Table 2. CBEEEs had no effect on expected years of schooling. The only variable that had a significant effect on how long young people typically stay in school was the nation's income.

### International Assessment of Educational Progress (IAEP) for Nine Canadian Provinces

When the Educational Testing Service canvassed countries about participating in the 1991 IAEP, Canada decided to collect sufficient data to allow reliable comparisons among provinces and between the Anglophone and Francophone school systems of the five provinces with dual systems.8 At the time, Alberta, British Columbia, Newfoundland, Quebec, and Francophone New Brunswick had curriculum-based provincial examinations in English, French, mathematics, biology, chemistry, and physics during the senior year of high school. These exams accounted for 50 percent of that year's final grade in Alberta, Newfoundland, and Quebec and 40 percent in British Columbia. The other provinces did not have curriculum-based provincial external exit examinations in 1990-91. Ontario eliminated them in 1967, Manitoba in 1970, and Nova Scotia in 1972. Anglophone New Brunswick had provincial exams in language arts and mathematics, but exam grades were not reported on transcripts or counted in final course grades. Canadian provincial exams are mediumstakes, not high-stakes, tests. They influence grades, but passing the examination is not essential for graduation. Employers appear uninter-

<sup>&</sup>lt;sup>8</sup> All French-speaking schools in New Brunswick, Saskatchewan, and Manitoba were invited to participate. Stratified random samples of 105 to 128 secondary schools were selected from the French-speaking school systems of Ontario and Quebec and the English-speaking school systems in all provinces, with the exception of Prince Edward Island.

	Hy- poth- e- sized sign (1)	Mean (2)	School Standard . Deviation (3)	Curricu Based Coeff. t	ulum- Exam t-stat. (5)	French speaking (6)	Religious School Board (7)	Log Books in Home (8)	Adjusted R <sup>2</sup> (9)
Achievement	( )	( )	(-)	( )	(-7	(-)	( )	(-)	(-7
Mathematics	+	470	135	051	(7.6)	07/***	- 0/8***	1/5***	320
Science	+	.541	.096	.026	(7.0)	.074	036***	.145	.323
Discipline Problems	0/+	.765	.720	017	(.4)	.19***	132**	282***	.080
Absenteeism Problems	0/+	.822	.766	.140	(3.1)	16**	.001	411***	.131
School Administra	tor Be	ehavior							
Math Specialist Teachers	+	.45	.50	.18	(6.9)	.08**	195***	.074**	.280
Science Specialist Teachers	+	.46	.50	.15	(5.6)	03	103***	.141***	.279
Took Math Courses in University	+	.64	.39	.19	(7.0)	06*	120***	.067**	.127
Took Science Courses in University	+	69	38	19	(8.5)	- 21***	- 172***	047	199
Math Class Hours	+	3.98	.88	.33	(5.9)	.31***	057	254***	.124
Science Class Hours	+	2.93	.79	.16	(3.5)	06	365***	006	.132
Computers per Student	?	.051	.043	.001	(.6)	006*	009***	.004	.195
Specialized Science Labs	+	1.95	.95	.28	(5.6)	.043	097	.037	.274
Teacher Behavior									
Total Homework Hours per Week	+	4.41	1.62	.65	(6.9)	48***	.621***	146	.149
Math Homework Hours per		1.66	64	01	(5.0)	00	100***	017	051
Science Homework Hours per	т	1.00	.04	.21	(0.0)	00	.109	.017	.001
Week	+	1.04	.47	.16	(5.1)	11**	.149***	.089**	.054
Math Quiz Index	+	1.62	.52	.10	(3.8)	.64***	107***	074**	.391
Science Quiz Index	+	.89	.38	.10	(4.9)	.32***	102***	007	.206
Home Behavior an	d Atti	tudes							
Average Hours of TV per Week in School		14 7	2 85	- 68	(4.2)	-1 7***	63*** -	-2 69***	255
Read for Fun Index	?	1.85	.28	.05	(2.8)	.08***	.028	.264***	.115
Watch Science Programs on	ว	07	38		(2.2)		.020	- 000***	001
	:	.01	.00	.00	(2.0)	. 4 1	.000	tinuod on	
							1001	nin lueu ol l	next page)

#### Table 3

## Effects of Curriculum-Based External Exit Exams in Canada

	Hy- poth- e- sized sign (1)	Mean (2)	School Standard - Deviation (3)	Curric Based Coeff. (4)	ulum- I Exam <i>t</i> -stat. (5)	- French speaking (6)	Religious School Board (7)	Log Books in Home (8)	Adjusted R <sup>2</sup> (9)
Parents Talk about Math Class	+	.62	.17	.04	(3.4)	.02	.044***	.016	.046
Parents Talk about Science Class	+	.47	.17	.06	(5.2)	01	.007	.074***	.056
Parents Want Me to Do Well in Math	+	3.54	.22	.05	(3.1)	01	.093***	.084***	.104
Parents' Interest in Science (0–4)	+	2.18	.34	.06	(2.6)	.12***	.109***	.209***	.071
Science Useful in Everyday Life	+	2.46	.31	.06	(2.7)	.18***	.141***	097***	.095

#### Table 3 (continued)

Effects of Curriculum-Based External Exit Exams in Canada

Note: On a two-tail test, \*\*\* indicates p < 0.01, \*\* indicates p < 0.05, and \* indicates p < 0.10. Controls also included mean number of siblings, the proportion of the students who use a different language at home, the number of students in a grade, and dummy variables for independent schools, religious schools, K–11 schools, and schools including 4<sup>th</sup> grade.

Source: Author's regressions predicting the characteristics of 1,309 to 1,338 Canadian secondary schools.

ested in exam scores. Job application forms do not ask applicants to report exam scores or grades.

The principals of schools sampled by IAEP completed questionnaires describing school policies, school resources, and the qualifications of eighth-grade mathematics and science teachers. Students were asked about books in the home, number of siblings, language spoken at home, hours of TV, hours doing homework, pleasure reading, watching science programs on TV, parental oversight of schoolwork, and teaching methods of teachers.

The effects of curriculum-based provincial exit exams taken by twelfth graders on the achievement and behavior of Canadian 13-yearolds, their parents, teachers, and school administrators were examined by estimating models predicting these behaviors using schools as observations. The data set comprises 1,338 Canadian schools. The model uses 11 explanatory variables: logarithm of the mean number of books in the home; the mean number of siblings; the proportion of the school's students whose home language was different from the language of instruction; logarithm of the number of students per grade in the school; dummies for schools run by a locally elected religious school board, independent secular schools, independent nonsecular schools, schools with primary grades, schools that include all grades in one building, and French-speaking schools; and a dummy for province exam.

Altogether, regression analysis was performed for four achievement outcomes, 12 measures of school administrator behavior, eight teacher behaviors, and 11 student/parent attitudes and behaviors. Table 3 presents the results for each achievement measure and a representative subset of the other variables of interest.<sup>9</sup> The first column presents the hypothesized sign of the relationship between CBEEE systems and that variable. The means and standard deviations across schools of each dependent variable are presented in columns two and three. The coefficient for the CBEEE dummy variable and its t-statistic are presented in columns four and five. The R<sup>2</sup> corrected for degrees of freedom is reported in the last column.

Provincial exit exams had large positive effects on student achievement: 19 percent of a U.S. standard deviation (about four-fifths of a U.S. grade-level equivalent) in mathematics and 13 percent of a standard deviation (about one-half of a grade-level equivalent) in science.

Exit exams also influenced the behavior of parents, students, teachers, and school administrators in Canadian provinces. Schools in exitexam provinces scheduled significantly more hours of math and science instruction, assigned more homework, had better science labs, were significantly more likely to use specialist teachers for math and science, and were more likely to hire math and science teachers who had studied the subject in college. Eighth-grade teachers in exam provinces gave tests and quizzes more frequently. The following were not significantly affected by CBEEEs: hours in the school year, library books per student, class size, or teacher preparation time (results not shown).

Opponents of externally set curriculum-based examinations predict that they will cause students to avoid learning activities that do not enhance exam scores. This hypothesis was examined by seeing whether exam systems were associated with *less* reading for pleasure and *less* watching of science programs like NOVA and Nature. Neither of these relationships was found. Indeed, students in exam provinces spent significantly more time reading for pleasure and more time watching science programs on TV, while watching significantly less TV overall. Parents in these provinces were more likely to talk to their children about their math and science classes, and their children were more likely to report that their parents "are interested in science" or "want me to do well in math."

Do CBEEEs skew teaching in undesirable ways? Apparently not. Students did more (not fewer) experiments in science class, and emphasis on computation using whole numbers—a skill that should be learned by

<sup>&</sup>lt;sup>9</sup> The remaining regression results are available from the author upon request.

the end of fifth grade—declined significantly (these results are not presented in the table). Apparently, teachers subjected to the subtle pressure of a provincial exam four years in the future adopt strategies that are conventionally viewed as "best practices," not strategies designed to maximize scores on multiple-choice tests.

Signed to maximize scores on multiple-choice tests. Students responded to the improved teaching by becoming more likely to report that science was "useful in everyday life." The data provided no support for our hypothesis that CBEEEs would induce employers to pay greater attention to high school achievement. Students in exam provinces were not more likely to believe that math was important in getting a good job and were less likely to believe that science was important in job hunting (results not shown).

One possible skeptical response to these findings is to point out that the correlation between the exam and other outcomes may not be causal. Maybe the people of Alberta, British Columbia, Newfoundland, Quebec, and Francophone New Brunswick—the provinces with exam systems place higher priority on education than do people in the rest of the nation. Maybe this trait also results in greater political support for examination systems. If so, we would expect schools in the exam provinces to be better than schools in other provinces along other dimensions, such as discipline and absenteeism, and not just by academic criteria. Bishop (1996) predicts, to the contrary, that exam systems induce

Bishop (1996) predicts, to the contrary, that exam systems induce students and schools to *redirect* resources and attention to the learning and teaching of exam subjects and away from the achievement of other goals such as low absenteeism, good discipline, and lots of computers. These competing hypotheses are evaluated in the third, fourth, and eleventh rows of Table 3. Contrary to the "provincial taste for education" hypothesis, principals in exam provinces had not purchased additional computers, did not report significantly fewer discipline problems, and were significantly more likely to report absenteeism problems.

### Scholastic Aptitude Test (SAT) in New York State

In the early 1990s, New York state was the only state with a voluntary CBEEE system. In 1993, about 56 percent of ninth graders took the mathematics course 1 exam and, of these, 24 percent of those not taking Regents exams were typically in courses that were considerably less challenging than Regents-level courses. A system of minimum competency tests in specific subjects set a minimum standard for those not taking Regents courses but, as in other states, the passing standard was low.

New York's students are more disadvantaged, more heavily minority, and more likely to be foreign-born than students in most other states. Among northern states, only Maryland, Delaware, and Illinois have a larger share of African-American pupils. Nationwide, only California has

	Basic Model	Teacher-Pupil Ratio and Spending per Pupil
New York State Dummy	46**	35*
	(2.7)	(2.0)
SAT Participation Rate	-68**	-88***
·	(2.6)	(3.3)
Type of Test-Taking Population:	· · ·	
Parents AA-BA+	370***	367***
	(6.4)	(6.6)
Private School	60	69*
	(1.6)	(1.9)
Black	-135***	-113
	(3.2)	(2.6)
Large School	-44*	-36
	(1.8)	(1.5)
Three or More Math Courses	85	45
	(1.3)	(.7)
Three or More English Courses	-36	-45
	(.3)	(.4)
 R <sup>2</sup>	.926	.933
RMSE	14.8	14.2

### Table 4

Determinants of Mean Total SAT-I Scores for States

Note: Numbers in parentheses are *t*-statistics. On a two-tail test, \*\*\* indicates p < 0.01, \*\* indicates p < 0.05, and \* indicates p < 0.10.

Source: Author's calculations.

a higher share of foreign-born population, and only California, Texas, Arizona, New Mexico, and Colorado have larger Hispanic population shares. In New York, literacy levels among adults are substantially below the national average (National Education Goals Panel 1993).

Consequently, when we compare student achievement levels, family background must be taken into account. Considering the high incidence of at-risk children, New York students do remarkably well. The proportion of students taking algebra, calculus, chemistry, and physics is generally above the national average. A larger proportion (9.4 percent) of New York's eleventh and twelfth graders are taking and passing AP exams in English, science, math, or history than any other state except Utah (National Education Goals Panel 1993).

Graham and Husted's (1993) analysis of SAT test scores in the 37 states with reasonably large test-taking populations found that New York state students did better than comparable students in other states. They did not, however, test the statistical significance of the New York state (NYS) effect and used an unusual log-log specification.

Table 4 presents the results of a linear regression predicting 1991

mean SAT-Math plus SAT-Verbal test scores for the 37 states for which data are available. With the exception of the dummy variable for New York state, all right-hand-side variables are proportions—generally the share of the test-taking population with the characteristic described. Clearly, New Yorkers do significantly better on the SAT than students of the same race and social background living in other states (row one).

When this model is estimated without the NYS dummy variable, New York has the largest positive residual in the sample. The next largest positive residual (Wisconsin's) is 87 percent of New York's residual. Illinois and Nevada have positive residuals that are about 58 percent of New York's value. Arizona, California, Colorado, Florida, New Mexico, Ohio, Rhode Island, Texas, and Washington have negative residuals greater than 10 points. Many of these states have large populations of Hispanics and recent immigrants, a trait that was not controlled for in the analysis. This makes New York's achievement all the more remarkable when one considers that Hispanics and immigrants are a large share of its schoolchildren.

For individuals, the summed SAT-Math plus SAT-Verbal has a standard deviation of approximately 200 points. Consequently, the differential between New York state's SAT mean and the prediction for New York (based on outcomes in the other 36 states) is about 20 percent of a standard deviation or about three-quarters of a grade-level equivalent.

Adding the teacher-pupil ratio and spending per pupil to the model reduces the NYS coefficient by 25 percent (column two). It remains significantly greater than zero, however. The teacher-pupil ratio has a significant positive effect on SAT scores. This suggests that heavy investment in K–12 schooling in New York state (possibly stimulated in part by the Regents exam system) may be one of the reasons why New York state students perform better than comparable students in other states.

The theory predicts that the existence of CBEEE systems will induce New York state to spend more on K–12 education and focus that spending on instruction. Indeed, New York's ratio of K–12 teacher salaries to college faculty salaries is significantly above average. New York teachers are also more likely to have master's degrees than are the teachers of any other state, except Connecticut and Indiana. New York ranks seventh in both teacher-pupil ratio and the ratio of per pupil spending to gross state product per capita (Bishop 1996).

Clearly, New York invests a great deal in its K–12 education system. If the cause of the high spending were a strong general commitment to education or legislative profligacy, we would expect spending to be high on both K–12 and higher education. This is not the case. New York is number one among the 50 states in the ratio of K–12 spending per pupil to higher education spending per college student.

The Regents exams have been low- to medium-stakes tests, not high-stakes tests. Exam grades counted for less than a quarter of the final grade in the course and influenced only the type of diploma received. Employers ignored exam results when making hiring decisions. Students were aware that they could avoid Regents courses and still go to college. Indeed some perceived an advantage to avoiding them:

My counselor wanted me to take Regents history, and I did for a while. But it was pretty hard, and the teacher moved fast. I switched to the other history, and I'm getting better grades. So my average will be better for college. Unless you are going to a college in the state, it doesn't really matter whether you get a Regents diploma (Ward 1994, p. 12).

Indeed, the small payoff to taking Regents exams may be one of the reasons why 40 to 50 percent of students elected to take watered-down local classes either to reduce their workload or to boost their GPA.

This has changed. In 1996, the Board of Regents announced that students entering ninth grade in 1996 had to take a new Regents English examination and pass it at the 55 percent level. The requirement to take and pass exams in five subjects applies to those entering ninth grade in 1999 or later. The English exam has become more challenging. The reading selections are longer and more difficult. The biggest change is that the exam is six hours rather than three, and students must write four long essays rather than two. One of the four essays asks for a response to two long literary passages that are presented to them for the first time. In January 2001, the prompt was:

Read the passages on the following pages [a memoir and an essay]. . . . Write a unified essay about the discovery of beauty. In your essay use ideas from *both* passages to establish a controlling idea about the discovery of beauty. Using evidence from *each* passage, develop your controlling idea and show how the author uses specific literary elements or techniques to convey that idea.

These prompts clearly call for deeper thinking about literature than the prompts used in past Regents exams. There is nothing rote or formulaic about teaching students how to handle essay questions like these. The pressures created by these exams are improving the teaching of literature and writing throughout the state. This is the true purpose of the Regents exam system.

## **CONCLUSIONS**

Our review of the evidence suggests that the claims by advocates of standards-based reform that curriculum-based external exit examinations significantly increase student achievement are probably correct. Students from countries with such systems outperform students from other countries at a comparable level of economic development on TIMSS-95, TIMSS-Repeat, PISA, and IEA reading studies. School enrollment rates are not reduced by CBEEE systems. Not only did students from Canadian provinces with such systems know more science and mathematics than students in other provinces, but they also watched less television and talked with their parents more about schoolwork.

Furthermore, schools in provinces with external exams were more likely to employ specialist teachers of mathematics and science; hire math and science teachers who had studied the subject in college; have high-quality science laboratories; schedule extra hours of math and science instruction; assign more homework in math, science, and other subjects; have students do or watch experiments in science class; and schedule frequent tests in math and science classes.

When student demography was held constant, New York state, the only U.S. state with a CBEEE system in the early 1990s, did significantly better than other states on the SAT-I test. The pressures created by these exams are improving the teaching of literature and writing throughout the state. This is the true purpose of curriculum-based external exit exam systems.

#### References

- American Federation of Teachers. 1995. Setting Strong Standards: AFT's Criteria for Judging the Quality and Usefulness of Student Achievement Standards. Washington, DC: American Federation of Teachers.
  - -. 1996. Making Standards Matter: 1996. Washington, DC: American Federation of Teachers.
- Association of Secondary Teachers of Ireland. 1990. Information Sheet opposing changes in Examination Systems.
- Beaton, Albert, et al. 1996. Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

-. 1996. Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

- Bishop, John. 1990. "Productivity Consequences of What Is Learned in High School." Journal of Curriculum Studies 22 (2): 101-26.
  - -. 1992. "The Impact of Academic Competencies on Wages, Unemployment and Job Performance." Carnegie/Rochester Conference Series on Public Policy 37 (December): 127-95.
  - -. 1996. "The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning." International Journal of Education Research 23 (8): 653-752.
- Bishop, John, Joan Moriarty, and Ferran Mane. 1997. "Diplomas for Learning, Not Seat Time: The Effects of New York's Regents Examinations." Paper presented at the Regents Forum in Albany, New York (October). Brooks-Cooper, Candace. 1993. "The Effect of Financial Incentives on the Standardized Test
- Performance of High School Students." Cornell University, master's thesis (August).
- Cameron, Stephen V. and James J. Heckman. 1991. "The Nonequivalence of High School Equivalents." NBER Working Paper No. 3804 (August).
- Chubb, John and Terry Moe. 1990. Politics, Markets, and America's Schools. Washington, DC: Brookings Institution.

- College Board. 1999. "More Schools, Teachers and Students Accept the AP Challenge in 1998–99." August 31. New York: The College Board.
- Competitiveness Policy Council. 1993. *Reports of the Subcouncils*. March. Washington, DC: Competitiveness Policy Council.
- Costrell, Robert. 1994. "A Simple Model of Educational Standards." American Economic Review 84 (4): 956–71.
- Edwards, Virginia B. 1999. "Quality Counts '99: Rewarding Results, Punishing Failures." Education Week 18 (17): 87–93.
- Fleming, M. and B. Chambers. 1983. "Teacher-Made Tests: Windows on the Classroom." In Testing in the Schools: New Directions for Testing and Measurement. San Francisco: Jossey Bass.
- Graham, Amy and Thomas Husted. 1993. "Understanding State Variation in SAT Scores." Economics of Education 12 (3): 197–202.
- Gummere, Richard. 1943. "The Independent School and the Post War World." Independent School Bulletin 4 (April).
- Hayward, Ed. 2001. "Dramatic Improvement in MCAS scores." Boston Herald (October 16).
- Hollenbeck, Kevin and Bruce Smith. 1984. Selecting Young Workers: The Influence of Applicants' Education and Skills on Employability Assessments by Employers. Columbus, OH: National Center for Research in Vocational Education, Ohio State University.
- International Assessment of Educational Progress. 1992. *IAEP Technical Report*. Vol. 1. Princeton, NJ: Educational Testing Service.
- Kang, Suk. 1985. "A Formal Model of School Reward Systems," in *Incentives, Learning, and Employability*, edited by J. Bishop. Columbus, OH: National Center for Research in Vocational Education, Ohio State University.
- Kiplinger, Vonda and Robert Linn. 1993. "Raising the Stakes of Test Administration: The Impact on Student Performance on NAEP." Center for the Study of Evaluation Technical Report 360. National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.
- Lerner, Barbara. 1990. "Good News about American Education." Commentary 91 (3): 19-25.
- Madeus, George. 1991. "The Effects of Important Tests on Students: Implications for a National Examination or System of Examinations." Paper presented at the American Educational Research Association Invitational Conference on Accountability as a State Reform Instrument: Impact on Teaching, Learning, Minority Issues, and Incentives for Improvement, Washington, DC (June).
- Madeus, George and Thomas Kellaghan. 1991. "Examination Systems in the European Community: Implications for a National Examination System in the United States." Contractor Report for the Office of Technology Assessment, U.S. Congress, Washington, DC.
- Merten, Don. 1996. "Visibility and Vulnerability: Responses to Rejection by Nonaggressive Junior High School Boys." *Journal of Early Adolescence* 16 (1): 5–26.
- Mullis, Ina, et al. 1997. Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- National Education Goals Panel. 1993. National Education Goals Report 1993. Vol. 2. Washington, DC: Government Printing Office.
- ——. 1995. Data for the National Education Goals Report: 1995. Vol. 1. Washington, DC: Government Printing Office.
- O'Neil, Harold F., et al. 1997. "Final Report of Experimental Studies on Motivation and NAEP Test Performance." Center for the Study of Evaluation Technical Report 427. National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.
- Organization for Economic Cooperation and Development. Center for Educational Research and Innovation. 2001. *Education at a Glance 2001*. Paris: Organization for Economic Cooperation and Development.
- Peter D. Hart Research Associates. 1995. Valuable Views: A Public Opinion Research Report on the Views of AFT Teachers on Professional Issues. Washington, DC: American Federation of Teachers.
- Rohwer, William D. and John W. Thomas. 1987. "Domain Specific Knowledge, Cognitive Strategies, and Impediments to Educational Reform," in *Cognitive Strategy Research*, edited by M. Pressley. New York: Springer-Verlag.
- Steinberg, Laurence, Bradford Brown, and Sanford Dornbusch. 1996. Beyond the Classroom. New York: Simon and Schuster.
- Thomas, John W. 1991. "Expectations and Effort: Course Demands, Students' Study Practices, and Academic Achievement." Paper presented at the Office of Educational Research and Improvement Conference on Student Motivation.
- Tinkelman, Sherman N. 1966. "Regents Examinations in New York State after 100 Years." Albany, NY: The University of the State of New York and the State Education Department.
- U.S. Department of Education. National Center for Education Statistics. 1993. *Digest of Education Statistics: 1992.* Washington, DC.
  - —. 1996. Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context: Initial Findings From The Third International Mathematics and Science Study. NCES 97–198. Washington, DC.
  - —. 2001. Outcomes of Learning: Results From the 2000 Program for International Student Assessment of 15-Year-Olds in Reading, Mathematics, and Science Literacy. NCES 2002–115, by M. Lemke, et al. Washington, DC.
- U.S. General Accounting Office. 1993. Educational Testing: The Canadian Experience with Standards, Examinations, and Assessments, by K. D. White. GAO/PEMD-93-11. Washington, DC.
- Ward, Deborah Hormell. 1994. "A Day in the Life." New York Teacher 25 (10): 10-12.
- World Bank. 2001. The World Development Report 2000–2001: Attacking Poverty. New York: Oxford University Press.

Appendix Table 1	
Examples of End-of-Course (EOC) Examination Systems	

State	Year Announced	Subjects (year first administered)	Score on Transcript	Part of Course Grade	Teachers Grade Exam	Honors Diploma Based on EOC Exam	Year Minimum Competency Exam (MCE) Begins	EOC Exam can substitute for MCE	Other Rewards for Student Achievement
NY	1865	English, Math, Biology, Chemistry, Physics, U.S. History, World History, Latin, Foreign Languages, Introduction to Occupations	Yes	Yes	Yes	Yes (40%)	1979	about 1992	In 1950s scholarships were based on Regents exams. Use in teacher assessment is a local option. Becomes primary high school graduation test after 2000–03.
NC	1984	Algebra I, Biology (1987); Algebra II, U.S. History (1988); Chemistry, Geometry (1989); English I, Physics, Social Studies (1990–91)	Yes	Most (25% after 2000)	Yes	2003	1980	No	State tests at earlier grades influence retention decisions.
CA	1983	Algebra I, Geometry (1987); U.S. History, Economics (1990); Biology, Chemistry (1991); Coordinated Science (1994); Writing (1996); Civics (1997); Literature, High School Math (1998); Physics, Spanish (1999)	Yes	No	No	Yes (1%)	2004	No	State tests at earlier grades influence retention decisions.
ТХ	1992	Biology (1995); Algebra I (1996); U.S. History, English (1999)	Yes	Most (required in future)	?	No	1987	2000	Scholarships based on course rigor and family income. State tests at earlier grades influence retention decisions.
TN	1992	Algebra I, Biology, English II (2001); Algebra II, Geometry, English I (2002); U.S. History, Chemistry, Physics (2003)	Yes	Yes	?	No	1985	2005	Becomes high school graduation test in 2005. Current honors diploma based on GPA.

276

## Appendix Table 1 (continued) Examples of End-of-Course (EOC) Examination Systems

State	Year Announced	Subjects (year first administered)	Score on Transcript	Part of Course Grade	Teachers Grade Exam	Honors Diploma Based on EOC Exam	Year Minimum Competency Exam (MCE) Begins	EOC Exam can substitute for MCE	Other Rewards for Student Achievement
MD	1995	English I, Civics, Algebra, Geometry, Biology (2001)	Yes	?	?	No	1983	2007	Becomes high school graduation test in 2007. Honors diploma based on rigorous courses and GPA since 1998.
MS	1994	Algebra, U.S. History (1997); Biology (1998)	?	?	?	No	1989	No	Merit Scholarship based on GPA and ACT scores.
VA	1996	English, Algebra I & II, Geometry, Earth Science, Biology, Chemistry, U.S. History, World History (1998)	Yes	Some	?	Yes	1981	2004	Becomes high school graduation test in 2004. State tests at earlier grades influence retention decisions.
OK	1999	English, U.S. History (2000); Math, Biology (2001)	Yes	No	No	No	None	No	State university and employers encouraged to use EOC exam in admission and hiring.
AR	1997	Math (1999); English (2002); Science, History (2004)	Yes	No	No	No	None	No	State tests at earlier grades influence retention decisions.

Appendix Table 2	
Examples of End-of-Grade (EOG) Examination Systems	

State	Year Announced	Subjects (year first administered)	Score on Transcript	Part of Grade	Teachers Grade Exam	Honors Diploma Based on EOG Exam	Year High School Graduation Test Requirement Begins	Other Rewards for Student Achievement
ОН	1987	12 <sup>th</sup> grade: Reading, Science, Math, Civics (1994–96)	Yes	No	No	Yes, in part	1994	\$500 scholarship based on EOG exam. Honors diploma based on rigorous courses, GPA, 12 <sup>th</sup> grade exams, or ACT.
CT	1991	10 <sup>th</sup> grade: English, Math, Science (1994)	Yes	No	No	Yes	None	
MI	1991	11 <sup>th</sup> grade: Math, Reading, Science, Writing (1997); Social Studies (1999)	Yes	No	No	Starts 1996, subject-by-subject	None	Beginning with 2000 graduates, \$2,500 scholarship awarded based on EOG exams.
PA	1991	11 <sup>th</sup> grade: Reading, Writing, Math (1999)	Some	No	No	Starts 2003	None	
OR	1991	10 <sup>th</sup> grade: English, Math (1996); Science (1999); Social Studies (2003)	Most; expect increase	Some	Teachers blind	Starts 2001 (proposed subject-by- subject)	None	Certification of Initial Mastery based on English and Math (2001), add Science (2002), Arts (2003), second language (2005), and Social Studies (2006).
IN	1993	10 <sup>th</sup> grade: English, Math (1997)	Most	No	No	No	2000	Graduation requirement also met by grade C or better in all Core 40 college prep courses or demonstrated 9 <sup>th</sup> grade achievement level. Honors diploma based on curriculum.
MA	1993	10 <sup>th</sup> grade: English, Math, Science (1998)	No, temporarily	No	No	Starts 2000	2003	As of March 2000, class of 2000 receive Certificate of Mastery based on EOG, AP, or SAT II scores.
IL	1997	11 <sup>th</sup> grade: Reading, Writing, Math, Science, Social Science (2001)	Yes	No	No	Starts 2002, subject-by-subject	None	
WI	1997	10 <sup>th</sup> grade: Reading, Writing, Math, Science, Social Science (2002)	Yes	No	No	No	2004	1997 high school graduation test legislation repealed in 1999; left to local districts.

Source: Author's research using multiple reference materials.

## Discussion

# WHAT IS THE APPROPRIATE ROLE FOR STUDENT ACHIEVEMENT STANDARDS?

## David N. Figlio\*

Politicians and policymakers across the political spectrum advocate high standards as a means of evaluating students. However, there exists very little published evidence that student achievement standards, such as the curriculum-based external exit exam (CBEEE) systems described by Bishop, lead to the substantial performance gains that advocates argue should occur. His paper takes an ambitious step in this direction. Bishop describes compellingly how a system of standards could change the culture of a school and its student body, then presents a series of empirical exercises in which he shows that, in cross-section, countries with CBEEE systems have higher performance than do countries without these systems. Canadian provinces with CBEEEs also do better than those without, and New York State, with its Regents exams, has a higher performance than might be predicted in the absence of its CBEEEs. Of these analyses, the New York analysis is the least plausible, because it relies on the presumption that New York is observationally equivalent to the rest of the country (holding observables equal) save for the Regents Exam, and the Canadian analyses are by far the most believable because the assumption of ceteris paribus is most likely met with detailed background characteristics controlled for and a single national educational system and infrastructure controlled for.

Each of these analyses points in the direction of standards leading to substantial improvements in student test scores in the tested subjects, and the Canadian analysis presents evidence of a series of mechanisms through which these standards might work. For instance, Bishop shows that schools in Canadian provinces with CBEEEs tend to focus instruc-

<sup>\*</sup>Walter Matherly Professor of Economics, University of Florida.

tional resources and homework time on the evaluated subjects (suggesting that schools respond to the standards) and that students and families in provinces with CBEEEs apparently alter their behavior, watching less television (but more science shows), talking more about coursework, reading for fun, and changing attitudes about mathematics and science, as examples of suggested behavioral modifications resulting from standards.

This research is consistent with the small amount of research in existence on the related topic of grading standards. Many of the mechanisms put forward by Bishop that might lead to improved student outcomes with CBEEEs also would lead to improved student outcomes when grading standards are elevated. Betts (1995) and Betts and Grogger (2000) present national cross-sectional evidence on the effects of schoollevel grading standards, measured by the difference between grade-point averages of students in the school and the same students' examination scores. These studies find that, on average, students perform better on examinations, attain more education, and earn more in the post-school early labor market when they attend high schools with high grading standards.

In my own research with Lucas (Figlio and Lucas 2000), we follow the same elementary school students over time and find that they learn more (measured by improvements in mathematics and reading scores) and behave better (measured by fewer serious disciplinary incidents) in a year in which they have a teacher with high grading standards than in a year in which they have a teacher with low grading standards. Both the papers by Betts and Grogger (2000) and by Figlio and Lucas (2000) demonstrate that there exist differential effects of grading standards on different types of students. Bishop's findings complement the results of this other literature nicely. Moreover, my paper on grading standards suggests that parents actually view high-standards teachers less favorably than they view teachers with softer grading standards, a finding consistent with Bishop's assertion that parents may view high standards (in the absence of a systemwide set of standards) unfavorably.

Taken together with these other results, Bishop's findings of very large, positive effects of CBEEEs on student outcomes might suggest that CBEEEs are a "silver bullet"—an educational intervention that dramatically improves performance at low cost. However, there are reasons to be skeptical of the magnitudes of the findings, if not the signs of the general relationships reported in this paper. The principal reason for concern is Bishop's identification strategy—that is, the way in which he empirically uncovers the relationship between CBEEEs and student outcomes. As it stands, it is difficult to be certain that these standards are driving the estimated results. All of Bishop's analyses rely exclusively on crosssectional variation. This type of variation is fine if there exist no omitted variables that might be correlated with both standards and student outcomes. But it is possible that some third variable could explain both standards and outcomes. One such case of this is that causality may be reversed, and countries (or provinces) where it is easier to meet standards, for unobserved reasons, are the governments that are most likely to impose them.

While Bishop alludes to this possibility, the issue is more substantial than its presentation in the paper. In his Canadian analysis, Bishop contends that there is little evidence that provinces with CBEEEs have higher tastes for education than those without CBEEEs. He provides evidence from Bishop (1996) suggesting that provinces with CBEEEs do not demonstrate increased tastes for education, as measured by improved discipline, attendance, or computer availability. I cannot speak to the computer access issue, but I find the results regarding discipline and attendance to be somewhat suspect. Using Florida data in the past, I have compared principal-reported measures of perceived discipline and attendance problems to actual discipline and attendance problems (as measured by administrative records). In these analyses, I have found that at best there is no correlation between principal perceptions of discipline and attendance problems and actual levels of discipline and attendance problems, and in most settings there is actually an inverse relationship between these measures. Principals in affluent schools may be more sensitive to these types of problems, and perceive even mild problems as severe, while principals in poor schools may perceive even serious problems as acceptable. (The same patterns are evident for drug problems, tardiness, teen pregnancy, and juvenile delinquency.) While Can-ada is obviously different from Florida, the conclusion drawn is that there is, at best, weak evidence against the presumption that provinces with CBEEE systems value education more.

But there is evidence, presented in the present paper, that seems to support the reverse causation argument. Some of the outcome variables discussed by Bishop may easily be thought of as causes of CBEEEs. For example, Bishop finds that parents in provinces with CBEEEs talk to their children more about math and science classes, and children in these provinces watch less television (but more science programming) and read for fun more. The conclusion drawn by Bishop is that these are outcomes of CBEEEs. This may certainly be the case. But it is just as likely, in my view, that these are attributes of the communities that impose CBEEEs, and thereby reflect tastes for education. While it is true that the provinces that imposed CBEEES run the gamut from the affluent west (Alberta and British Columbia) to more moderate Quebec to the poor provinces of New Brunswick (Francophone portion only) and Newfoundland, the population distribution of these provinces is such that the sample is dominated by Alberta, British Columbia, and metropolitan Quebec. In the 2001 Census, nearly three-quarters of the population of these provinces resided in Alberta, British Columbia, and the Montreal metropolitan area, implying that the population of the CBEEE provinces is not as diverse as one might expect given their numbers.

Bishop also presents cross-sectional evidence suggesting that provinces with CBEEE systems dedicate more resources to the topics covered by CBEEEs. For instance, schools in these provinces have more math and science specialist teachers, more math and science class hours, and teachers with more math and science experience. These results suggest that CBEEEs lead to institutional behavioral changes. However, in a cross-section, it is impossible to be certain of the direction of the causality. It may be that these variables indicate that the provinces ultimately imposing CBEEEs have a greater taste for mathematics and science instruction, tastes that are reflected both in curricular emphasis and in standards-setting. If this latter explanation is true, then it may be the case that the differences in resource use (and, presumably, in outcomes) would have existed in the absence of CBEEEs. We have no way of distinguishing these two explanations, and, therefore, the paper would be strengthened considerably if some within-school, over-time variation could be exploited. While Bishop's results are plausible and compelling, they are not fully convincing, and will never be so unless one can be more certain that the identification problem is solved. This will not occur in a crosssectional setting. Ultimately, while I believe that CBEEEs lead to higher average performance, I do not know whether the magnitudes put forward by Bishop are accurate.

Bishop's analysis looks only at the mean effects of CBEEEs. This is a necessity in his cross-national analyses, but is not necessary in the case of his Canadian research. More research needs to be done to look at the distributional consequences of CBEEEs. This is important for several reasons. First, the existing theoretical research on standards, including work by Betts (1998) and Costrell (1994), suggests that they might have differential effects on students at varying parts of the ability distribution. One can tell stories in which high achievers and low achievers could either be helped or harmed by a CBEEE. Bishop presents arguments for how these students could be helped. But high achievers who are likely to have exceeded the standard without additional effort may work less, and low achievers who are unlikely to make the standard under most circumstances may give up and work less as well. (For instance, Lillard and DeCicca, in their forthcoming article, find that high graduation coursework standards induce greater dropout rates.) I have no way of ruling out these possibilities, so the conflicting stories that can reasonably be told make the question of the distributional consequences of CBEEEs an empirical one. While the research on grading standards mentioned above suggests that few, if any, students are harmed by high grading standards, there is still evidence that high achievers may benefit more. I hope that Bishop, in his future work, will investigate whether CBEEEs

help certain types of students, or students in certain types of settings, more than other types of students.

Regardless of the magnitude of the effect of CBEEEs, the present policy environment may present challenges for their implementation. Dozens of states currently have test-based systems of school accountability—effectively, high- or medium-stakes standards for schools—and with the federal No Child Left Behind Act of 2001 these stakes are elevated nationwide. On July 1, 2002, the U.S. Department of Education deemed 8,652 Title I schools nationwide sufficiently in need of improvement that students attending them are eligible for enhanced public school choice. In most states, schools and students are evaluated on the same curriculumbased test. For instance, the Florida Comprehensive Assessment Test determines not only school rewards and sanctions, including eligibility for private school vouchers, but also student promotion in Florida. High student performance on the FCAT both ensures student promotion and helps schools earn a higher performance grade, with financial and governance ramifications for the school.

School accountability systems that evaluate schools using the same curriculum-based examination used to evaluate students may have the effect of setting student standards lower than what might ordinarily have been set. This may be even more the case under the No Child Left Behind law, where school "failure" is tied to removal of federal dollars, and hence states may prefer to sanction fewer schools than they might have in the absence of the federal law. Even before the passage of No Child Left Behind, the state of Florida postponed its planned increases in student standards in a move that was publicly speculated at the time of decision (though there is no definitive evidence of this) to be caused in part by the implications for Florida's own school accountability system. The incentives are much clearer toward setting low proficiency standards under the new federal law.

If student standards are set very low, however, one might ask whether low standards are better than no standards at all. An analogy might be made with teacher merit pay. In a recent working paper, Lawrence Kenny and I (Figlio and Kenny 2001) suggest that student performance is lower in schools that give merit pay to all or most teachers, regardless of teacher productivity, than if no merit pay is offered at all. On the other hand, offering merit pay to a small fraction of teachers tends to increase student test scores substantially in our U.S. national sample of students in schools. While this is by no means definitive, it is suggestive that low standards might be less productive than no standards. This leads one to ask whether students and schools should be evaluated on the same standard. The twin goals of student and school accountability may be met more easily if the two are uncoupled.

Bishop is to be commended for the work that he has done in assembling evidence on the effects of CBEEEs from so many different sources and in so many different settings. His work is provocative and extremely interesting, empirical identification issues notwithstanding. His title, "What Is the Appropriate Role for Student Achievement Standards?" is a relevant question to ask with respect to standards for schools as well. Is it appropriate to use student achievement standards to evaluate schools? And if not, one must ask, If schools are evaluated on a low standard, and student standards are multi-level, to whom will the schools pay attention? The answers to these questions are difficult to know right now with the current research, but the questions must be asked as the nation embarks on its new experiments with student and school accountability.

#### References

- Betts, Julian. 1995. "Do Grading Standards Affect the Incentive to Learn?" University of California, San Diego, working paper.
- ———. 1998. "The Impact of Educational Standards on the Level and Distribution of Earnings." American Economic Review 88 (1): 266–75.
- Betts, Julian and Jeff Grogger. 2000. "The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings." NBER Working Paper No. 7875 (September).
- Bishop, John. 1996. "The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning." International Journal of Education Research 23 (8): 653–752.
- Costrell, Robert. 1994. "A Simple Model of Educational Standards." American Economic Review 84 (4): 956–71.
- Figlio, David and Lawrence Kenny. 2001. "Does Teacher Merit Pay Improve Student Test Scores?" University of Florida, working paper.
- Figlio, David and Maurice Lucas. 2000. "Do High Grading Standards Affect Student Performance?" NBER Working Paper No. 7985 (October).
- Lillard, Dean and Philip DeCicca. Forthcoming. "Higher Standards, More Dropouts? Evidence Within and Across Time." *Economics of Education Review*.

## Discussion

# WHAT IS THE APPROPRIATE ROLE FOR STUDENT ACHIEVEMENT STANDARDS?

## Ellen Guiney\*

The Boston Plan for Excellence, a private, nonprofit organization, is a local education fund whose mission is to improve instruction and student performance in the Boston Public Schools. As such, we often work with and learn from education reform taking place in the nation's largest cities so as to bring lessons to Boston from other districts. Our knowledge base is derived primarily from studying what is happening to the students in the 35 largest cities. These students comprise 15 percent of the country's students. Generally, these students are poor, and they often begin school without preschool or other advantages enjoyed by middleclass children. A majority of them are children of color. A few statistics from our experiences are relevant:

- In a study of three Boston kindergarten classes that tested students' skills upon entering, Voices of Love and Freedom found that 60 percent of students knew fewer than 10 capital letters, 70 percent knew fewer than 10 lower-case letters, and 90 percent could make fewer than 10 letter-sound correlations.
- Nationwide, only 68 percent of all students complete high school in four years; in the 35 largest cities, fewer than 50 percent do so.
- Nationwide, half of ninth graders entering high school read at a sixth grade level.

## **Response to Bishop**

Our experience in Boston coincides with Bishop's conclusions. Our on-site observations in 50 Boston and other schools is that curriculum-

<sup>\*</sup>Executive Director, Boston Plan for Excellence.

based exit examinations do what he suggests. Tests that are aligned with and carefully measure high standards do affect a school's priorities, teachers' decisions, and students' decisions; they also influence the redirection of resources within schools to core subjects.

If we are to meet our civic and moral responsibilities as a country, however, setting standards, aligning assessments to measure whether students are learning them, and creating an aligned accountability system are only the foundation. Without the creation of a coherent system of improving instruction in classrooms—which will involve extensive professional development for principals and teachers and a deliberate reorganization of schools—urban students will not meet standards.

#### **EXAMINATION OF UNDERLYING ASSUMPTIONS**

Standards-based reform rests on certain assumptions that do not hold in the 35 largest cities. These assumptions include the following:

- 1. The system described will give students and teachers information about what students are not learning in a timely, usable way;
- 2. Students will be motivated to invest more in their learning because they face consequences and because they realize how much they need to know;
- 3. We have the right type of classroom instruction for these students;
- 4. There is an adult accountability system that creates the right information, sanctions, and incentives that lead to instructional improvement;
- 5. There are other professionals who are better equipped, prepared, and willing to take the places of those let go because they do not succeed with students; and finally,
- 6. Schools are coherently organized, at scale, to respond to the standards-based foundation laid out by Bishop.

Such an accountability system rests on the idea that educators already possess all the knowledge and skills they need to bring about substantial improvements in instruction that will lead to greater learning for urban students. Further, it assumes that teachers and principals (a) need more political and civic pressure to do what is effective; and (b) are not rewarded enough by the present system to be motivated to do what they know they should.

Let's look more closely at each of these assumptions.

Assumption 1. The present system gives good information that helps teachers know the extent and depth of student learning in a timely way. Virtually no large city district has a data system that puts into the hands of teachers and principals fine-grained, user-friendly information about individual students that teachers can use on an ongoing basis. For the most part, large-scale testing directed by the state takes place once a year, and considerable time passes before individual student results are reported. Students have usually moved on to a different grade and teacher, and sometimes to a different school. Few districts have a "formative" system to supplement these summative tests, and even if they do, management of these data is a challenge at the school level. Further, principals and teachers have not been trained in data analysis, so what might be useful lies unused.

Assumption 2. Students will be motivated to invest more time and energy because they face consequences, and because they are aware of how much they need to learn. Most teachers can teach the students easiest to teach (those who come into their classes with the store of basic knowledge about literacy and numeracy that readies them to learn). These students also come with the understanding that leads them to value what the teachers tell them is important, or at least they are compliant enough to suspend their disbelief.

Many of those far behind, however, have had previous academic experiences that have led them to believe school has little value for them. They may know abstractly that it has value but are not convinced that they will benefit from it because they never have. Many do not read well enough to learn or to enjoy reading, for instance. They know that they face consequences, and are disappointed in themselves, but have no sense of how they might turn the situation around.

Assumption 3. We have the right type of instruction for these students. Most teachers have not been prepared to teach students with differing levels of preparedness and knowledge, nor do they work in schools organized to make differentiated instruction reasonably possible. They do not know how to assess accurately where each student is, design a course of study for each depending on need, and then manage all these different levels. Most do not know how to teach in a sophisticated, highly intellectual way to build students' knowledge and skills, which is what a standards-based reform system requires.

There is a further problem with the instruction urban students receive. The crucial relationship in teaching is the one between the teacher and students, and their mutual engagement in the content. In most urban classrooms, however, teachers have an uneasy sense of the unknown and unknowable lives of their students and fear losing control. This leads teachers to minimize interactions with students and to make the exploration of content, ideas, and differences rare, even though these are essential to the higher learning demanded by standards. There is little "talk" or discourse in average urban classrooms. All of this results in a lacking sense of efficacy on either teachers' or students' parts, and little overall engagement. Finally, the organization and use of resources and time in most schools is not conducive to change or improvement. Adults have little time to learn or to interact with each other, nor a means to reorganize themselves.

Assumption 4. Adults are held accountable for the learning of students. In virtually all states and urban districts, the unit of analysis for accountability is the district, the school, or the student. It is not the teacher. Principals are evaluated, and are sometimes held accountable and let go if there are available replacements. But because replacements are not often available, districts stick with mediocre principals. Furthermore, the current teacher evaluation system rarely includes student performance results, teacher knowledge of the material on the state's standards, or the practice of effective pedagogy. When teachers are evaluated, the evaluation seldom includes an analysis of the effectiveness of their instructional practices in a deep way that leads to improvements in their classrooms. Many teachers report being visited by their principal rarely.

Intensive teacher evaluations are usually centered on the worst teachers, not the average ones, and rarely do evaluations highlight and elevate the superb practice of the best, who are obtaining wonderful performance with their students. Many critics blame unions for protecting teachers, but teacher evaluation problems go well beyond teachers' union issues. Although few districts have the contractual relationships right yet, there could be steps taken within existing contracts that would begin including student performance as part of evaluations. This would lead to a more robust adult accountability system.

Assumption 5. There is a supply of well-prepared and interested individuals ready and willing to step into urban classrooms, were we ready to terminate the mediocre ones. This is demonstrably untrue, as a look at California and cities elsewhere makes clear. Reports by the Education Trust, the National Commission on Teaching for America's Future, and others have documented the supply problem well. Beyond the numbers, even when states are tightening up qualifications, they tend to be assessing only low-level skills of future teachers. Teacher preparation institutions receive accreditation routinely without making any substantive changes in how teachers are prepared and trained.

Assumption 6. Schools and systems are coherent, and we have, at scale, examples of how to organize time, money, people, and support to get instructional improvement. We do not have the examples of high-performing districts that we need. The knowledge base about large-scale improvement is shallow. The Annenberg Institute for the Redesign of Urban School Districts has created a task force to find good models to

inform and improve support for schools so that instruction improves, but their work is incomplete. To date, they have found some high spots, but overall, the research in this field is weak.

### LOOKING TO THE FUTURE: IS THE SITUATION HOPELESS? NOT NECESSARILY

The situation can be changed if we collectively take several important steps.

**Step 1**: Recognize and accept that upping the stakes and consequences for schools and students, as the new federal legislation No Child Left Behind does, will not by itself cause instructional improvement, school coherence, or improved student performance.

**Step 2**: Conduct much more research on instructional improvement and then highlight the visible models of how it takes place. Many cities, like Boston, have parts of the answer, but San Diego, several New York community districts, Cincinnati, Long Beach, Houston, Denver, and many others have other pieces to the puzzle.

**Step 3:** Start making greater investments in the right things: improving teacher and principal knowledge about content, pedagogy, and the relationship between them. School staff cannot do what they do not know how to do at a high level: teach urban students to master challenging content no matter where they begin and how far behind they are.

**Step 4**: Get the data systems right so that they yield useful and fine-grained information for students, parents, and teachers. Technology has a greatly underdeveloped role in helping to solve this problem, but the knowledge and skill of teachers and principals to reflect on and use data about students' performance also must be addressed.

**Step 5**: States and the media should stop misusing assessments so we can build public understanding of the true problem and the solutions. Tests are important and useful, but not a good instrument to pinpoint the problem to be solved.

## Panel Discussion

## THEORIES OF ACTION FOR EFFECTING EDUCATION REFORM

Chester E. Finn, Jr.\*

This panel is meant to examine policies to improve educational outcomes, which I believe is everybody's objective. It is certainly what the country has been talking about for the last two decades, since it was declared *A Nation at Risk* in 1983. During this time we have had a lot of flailing about. We have tried a lot of things and have had a lot of false starts. There has been a great deal of activity, and we have done a lot of spending. We are still in the middle of this today. But we do not have much improved achievement to show for what we have been up to these past 19 years. You can find very spotty evidence, but earlier sessions have made clear that there is not a lot of conclusive evidence that learning outcomes have improved much.

Going forward, what should we do differently? I think it is useful to proceed with some kind of a theory of action, that is, some plausible notion of what we think is most apt to drive the improved results that we seek. Otherwise, we are bound for more flailing about. While it is a considerable over-simplification, I have found it useful to think in terms of four theories of action that I believe dominate the education-reform arena today. I think that two of these theories have some promise. Two of them do not, but we will nevertheless surely continue to use them. In the real world, we commonly find more than one of these theories operating at the same time in a given place. It gets very complicated when you start mixing and matching and coming up with hybrids. On the other hand, that is the real world, and I think we are probably going to discover at the end of the day that a hybrid will work better than any one of these strategies taken alone.

<sup>\*</sup>President, Thomas B. Fordham Foundation.

Of the four theories that I think dominated the conference discussion, two operate chiefly within the familiar framework of educational institutional arrangements and professions. The other two are driven mostly from outside the familiar institutional arrangements, and they include most of what we have been talking about at this conference. These outer-driven strategies are highly behaviorist. They presume that if you push from outside the system, people inside the system will begin to behave differently. Because they are highly behaviorist, I assume that they will be especially appealing to economists, just as they are deeply repugnant to most educators, not including myself.

The first theory of action is the oldest, the most familiar, and it is barely a changed system or strategy—just trusting the system to do a better job with additional resources. The assumption is that the school board and the superintendent want to do better, know how to do better, and would successfully improve student achievement, if only they had the wherewithal. This theory leads to a wide variety of resource-based strategies, such as smaller classes, longer school days, new textbooks, more technology, and so on.

The second theory of action, highly popular within the education field, I call, "trust the experts." The idea is that education experts, such as researchers, professors, gurus, and some of the people in this room, would know how to make the system work better, if only they had greater influence over it. Therefore, if we give them more influence, the results will improve. This seeks either to give more power to experts or to bring greater expertise to bear on what schools do. Examples of implementations of this strategy include installation in a school of a whole new comprehensive school design devised by somebody like Jim Comer, Mark Tucker, or Howard Gardner; or the introduction of matheducation experts to retrain fourth-grade teachers in a school so they can do a better job of teaching math. Warren Simmons and his colleagues are much involved with some of these kinds of activities under the Annenberg aegis and have been doing quite a lot of this for the last decade or so.

The third theory, which had most of the discussion at this conference because it is the most visible reform strategy in America today, I would call, "trust the government." It assumes that a state government, with a lot of "oomph" added by the federal government, will set standards, develop tests, and impose consequences on the education system, causing the people within the system to teach better, study harder, learn more, and so on. This strategy involves statewide academic standards and tests or assessments. If it is fully fledged, it also normally involves rewards and interventions or punishments, with the rewards going to those who meet the standards—be they students, schools, or educators—and with the interventions or punishments going to those who do not—to those who need to work harder or have their behavior changed in some way. You might hold the child back, dock the teacher's pay, replace the superintendent, or do some other intervention, or many other interventions, as cataloged with great precision year by year in the No Child Left Behind Act. The theory is that the government will create what a colleague called a kind of exoskeleton around this soft body of the education system and will thus cause it to shape up in a way that it would not do on its own.

The fourth strategy is, of course, competition and choice. This strategy also comes from outside the system, but it comes from the customers, from the marketplace, from, as it were, the bottom up rather than the top down. The theory, familiar to economists and others, is that the system will improve if it has competition; efficiency, quality, productivity, and performance will improve if there are choices, diversity, and marketplace forces at work. Theoretically this strategy will not only benefit the kids who are directly served by these alternative arrangements—typically low-income kids who otherwise would be trapped in failing, urban school systems—but will also tone up the whole system by virtue of the competition that is brought to bear upon it. This theory leads to charter schools, to vouchers, and to a myriad of other arrangements that come under the heading of competition and choice.

My evaluation of these four theories is as follows. I have very little confidence in trusting the system to do more with additional resources. There are occasional fluky situations where that strategy works, especially if really inspired leadership exists at a state, a district, or even a school building level. But these cases are rare. I also do not have a lot of confidence in trusting the experts to fix the system, though often expertise is needed within or in combination with one of these other strategies, as was discussed by Ellen Guiney, among others. The question, however, is whom do you trust to have the leverage to make things change? The experts are not the answer to that question. I do have a fair amount of confidence in the exoskeleton, the government-based, standards-based approach. But it is extremely difficult to implement successfully: to get the standards right, to get the tests right, to get the consequences in place, and so on. I also have a fair degree of optimism about the competition system, but its politics are so gnarly that we have not even given it a proper test, let alone given it a full-fledged endorsement as a reform strategy.

So where do we end up? Probably the most interesting example in front of us is the charter school phenomenon, which is a hybrid of strategies three and four. Charter schools combine accountability, tests, and information on the one hand, with diversity, competition, pluralism, and choice on the other hand. With 2,400 charter schools, heading rapidly toward 3,000, we have the beginnings of a naturally occurring experiment at the intersection of strategies three and four, and one that incidentally draws in a lot of expertise. A good charter school often has some inspired educators working on some very interesting ideas. It is simultaneously accountable both to the state for meeting the state standards as prescribed by the state test and to its customers. Because nobody has to attend them, charter schools will have no students—and no revenue—if nobody chooses to go to them.

Like Eric Hanushek, I would like to suggest that the standards-based, or government-driven, system and the market system might even need each other. Each of them by itself supplies a solution to the biggest problem besetting the other one. The market by itself lacks informed consumers, a problem that can be solved by government standards and tests. The government-driven accountability system is very good at identifying failing schools, but it is lousy at doing anything about them, for which the choice system may hold the solution. I have hopes about these two strategies working in combination as we go forward, and I think we are beginning to learn quite a lot about them from the charter school experience.

I used to assume that school inevitably meant a grown-up in a room with four walls and 20 little people. The most interesting thing that I am involved with right now, along with former education secretary Bill Bennett, is a private start-up called K–12 (K12.com) that is seeking to create a virtual school for the country and potentially for the world, with thousands of kids to be enrolled by September. When you begin to think through the implications of a virtual school, everything changes: the definition of a school, of a teacher, of a school day, of a learning environment, of what it means to be in third grade, and so on. I want to suggest that we might find ourselves, with the help of technology, actually catapulting over a lot of the institutional arrangements that have been so frustrating, exacerbating, and perplexing as we try to reform the system that we have today.

## Panel Discussion

## IMPROVING EDUCATION OUTCOMES: IN COLLEGES, UNIVERSITIES, AND BEYOND

## Alan G. Merten\*

While most of the attention of the public, the policymakers, and this conference is on improving our K–12 education system, our higher education also needs attention. Higher education is undergoing dramatic changes that have to be considered in conjunction with the demands on and the changes in the K–12 system. Similarly, there are common needs of the two systems, and lessons from higher education should be considered in any K–12 reform proposal.

Not that long ago, the vast majority of college undergraduates were 18 to 22 years old, lived on campus, and were full-time students. It is estimated that students with these attributes account for less than 20 percent of today's undergraduate population. In addition to their graduate education and research missions, higher education institutions have had to change to serve this very different population.

At all levels, the job of educational administrators has changed and will continue to change. Competent management of educational institutions by principals, superintendents, deans, and presidents is no longer sufficient. Just as in the corporate arena, bold leadership is necessary. Educational leaders must be willing and able to take risks, make decisions, and recover from and learn from failure. They must be eager to measure what is important. They must be aware of all of the sources and uses of funds available to them. They must eagerly accept responsibility and accountability.

In the midst of these tight economic times, the focus of society's attention concerning all elements of education at all levels has shifted from one of support and encouragement to one of cost-cutting and

<sup>\*</sup>President, George Mason University.

accountability. This shift of emphasis has occurred most notably in the eyes of elected officials and business leaders.

It is the responsibility of K–12 leaders and higher education leaders to reverse this trend. We need collectively to do a better job both of being effective and efficient managers of resources and in aggressively making the case for adequate resources. We need to make clear the connection between education and economic and social prosperity.

Higher education institutions have played a leadership role in providing the human capital that has created our successful information and knowledge-based economy. One element of the success story has been the increasing number of women, ethnic minorities, and non-U.S. citizens who have taken advantage of our higher education system and then contributed to our economy and our society, in general.

The workforce of the information society will need even more participation from these previously underrepresented groups. High school counselors and teachers will need to be better informed of career opportunities, and will need to direct the best and the brightest more proactively toward areas of national need and the potential for personal achievement. Since the tragedy of September 11, there have been proposals to limit severely the number of non-U.S. students in our colleges and universities. Many of these proposals have serious, harmful unintended consequences for our society. Education leaders need to play an active role in ensuring that our workforce draws upon all sources of talent, regardless of their country of origin.

Our system of higher education is the envy of the world. On the other hand, our K–12 system is apparently in need of reform. Are there lessons from higher education that may be applicable to K–12? There are three aspects of our higher education system that seem to be crucial. First, there is tremendous competition among the suppliers of higher education, and consequently, choice for the consumer. Second, merit-based compensation is a major component of the higher education and beyond) lead to compensation patterns that are based on academic discipline and areas of expertise. Public policymakers should assess what lessons can be learned from this for our K–12 system.

Finally, this conference has highlighted a range of economic analyses of various educational and related reforms. Because of the complexity of educational delivery systems and measures of educational outcomes, the research often has had to make simplifying assumptions. My plea to researchers in this area is to do all that is necessary to avoid the easy-to-study problems and concentrate on the important questions. Your audience, the populace, and the public policymakers cannot afford anything less.

## Panel Discussion

## IMPROVING URBAN PUBLIC SCHOOLS: SUGGESTIONS FOR TEACHER UNION LEADERS

## Richard J. Murnane\*

For three related reasons, I focus my comments on the challenge of improving the quality of education that urban public schools provide. First, the quantity and quality of education American children receive today has a much larger impact on their earnings prospects than was the case for American children 40 years ago. Second, minority children and children from low-income families have both lower cognitive skills and lower educational attainments than do white children and children from middle-class families. Third, 42 percent of individuals living in poverty in the United States live in central cities, and 40 percent of minority children attend school in one of the nation's largest 50 school districts.

In an earlier session of this conference, one participant asked the panelists what advice they would give to an audience of union leaders representing teachers from urban public schools. In response to that question, I offer seven suggestions. Following my explanations of those suggestions, I conclude with slightly different versions that I believe serve as relevant advice to local and state educational policymakers.

#### Suggestion One: Improve Student Outcomes or See the End of Public Schools as We Have Known Them

Since the passage of the first charter school legislation in Minnesota in 1991, the number of charter schools in the United States has increased to 2,400, enrolling more than 500,000 students. In June 2002, the U.S. Supreme Court ruled that allocating public funds to pay for educational

<sup>\*</sup> Thompson Professor of Education, Graduate School of Education, Harvard University. The author thanks Simone Sangster for first-rate research assistance.





Note: Vertical axis refers to percentile position of decade median of industry task input in overall 1959–98 distribution. Source: Revised from Autor, Levy, and Murnane (2001).

vouchers that low-income students in Cleveland use to pay for education in Catholic schools does not violate the U.S. Constitution. These are just two indications of changes in the political climate of the country in regard to publicly funded education. Unless urban public schools become more effective in educating children, the number of charter schools and voucher programs will increase, drawing students, money, and political power away from conventional public schools, and ending the promise of public schools as we have known it.

As representatives of the people who do the teaching in American public schools, teacher unions need to be involved in the design and implementation of improvement strategies. This is a necessary condition for the improvement of urban schools. At the same time, even the Herculean efforts of teachers cannot by themselves improve urban schools. Teachers must have the resources to do their jobs within an organizational structure that supports and demands excellence.

#### Suggestion Two: Insist on Meaningful Measures of Student Outcomes

Changes in the American economy over the last 30 years have dramatically altered the type of work Americans do and the skills they need to earn a decent living. As illustrated in Figure 1, there has been a dramatic decline in the proportion of the workforce engaged in routine cognitive tasks (for example, filing, bookkeeping) and an increase in the proportions engaged in activities that MIT economists David Autor and Frank Levy and I call expert thinking and complex communications. These changes are relevant to the design of the tests used to measure the skills of American students. If tests measure solely reading comprehension and the ability to do computations, these are the skills that will be emphasized in instruction. They are not the skills Americans need to earn a good living. Good teachers know this. A consequence is that the nation's ability to attract and retain effective teachers will require, among other things, that the tests used to measure student outcomes are worth teaching to.

The following writing prompt was part of the Massachusetts Comprehensive Assessment System (MCAS) exam, which was administered to 10th graders in April 2002 (the test high school students in the state must pass in order to receive a high school diploma):

In literature as in life, people struggle with principles or beliefs they hold. From a work of literature you have read in or out of school, select a character who struggles with his or her own principles or beliefs. In a well-developed composition, identify that character and explain how that character's inner struggle is important to the work of literature.

In my view, the state's investment in a rich, open-ended assessment system for measuring student skills is a step in the right direction toward the goal of improving education in the state.

## Suggestion Three: Demand a Level Playing Field with Charter and Voucher Schools

Increasingly, public schools are being judged against the performances of charter schools and voucher-supported private schools in educating students. For this to be a fair competition, the different types of schools should play by the same rules. Critical rules concern obligations to serve students with disabilities, students whose first language is not English, disruptive students, and mobile students. These students are relatively expensive to serve, and schools will volunteer to serve their share of such students only if they receive adequate compensation for doing so. If schools are unwilling to serve these groups of students, this is *per se* evidence that relative student funding levels are not set appropriately.

It is important to keep in mind that there is extraordinary uncertainty about the effectiveness of today's non-public schools, especially in educating urban children. More than 99 percent of the quantitative evidence documenting the relative effectiveness of private schools con-



cerns solely Catholic schools. The reason this matters is that Catholic schools serve a declining proportion of American students attending non-public schools. In 1960, Catholic schools served more than 90 percent of the students in non-public schools. The comparable percentage today is less than 45 percent (see Figure 2).

## Suggestion Four: Learn from Pilot Schools about How to Obtain Resources and Flexibility

Following the passage of the 1993 educational reform legislation that authorized the first charter schools in Massachusetts, the Boston Teachers Union (BTU) negotiated the creation of pilot schools in Boston. The teachers in these interesting public schools are BTU members who have agreed to waive certain elements of their contract regarding work rules. Currently Boston has 12 pilot schools. These schools receive lump-sum payments from the school district based on their student enrollments and have considerable flexibility in using their resources to design and carry out instructional programs. While they may use resources to buy services from the school district, they are not obligated to do so. The success of the pilot schools in attracting talented teachers and in developing interesting instructional programs suggests a promising strategy for developing schools staffed by teacher union members that can create innovative programs for educating urban children.

## Suggestion Five: Demand the Tools and the Time to Learn from Student Assessments

One of the strengths of the MCAS is that all questions that affect student scores are made public shortly after the tests are administered. Moreover, school faculties receive information on the performance of every child on every test score item. This creates the potential to learn a great deal from the assessment results about the skill deficiencies of individual students, about the weaknesses of instruction in particular schools and classrooms, and about how well schools are working in general. In fact, I would argue that the opportunities for learning about how well schools are working are at least as great as the strategies W. Edwards Deming advocated businesses adopt to learn about the effectiveness of their production processes.

Unfortunately, relatively few schools learn much from MCAS results. One reason is the delay in providing results to schools; a second is a lack of tools for analyzing the results efficiently; a third is a lack of training on how to do potentially powerful analyses; and a fourth is a lack of time in the school day to learn from the assessment results. If teachers are to be responsible for improving student achievement, their representatives should demand the tools and the time in the work schedule to learn from these potentially valuable assessment results.

#### Suggestion Six: Make Professional Development Work

Success in preparing urban students to pass demanding high-stakes exit exams like the MCAS requires changes in how teachers teach. Professional development is the term used among educators for the training aimed at improving how teachers teach and what students learn. Unfortunately, most professional development has little or no impact on how teachers teach.

In recent years, our understanding of which components of professional development improve instruction has increased. We now know that effective professional development must focus on teaching particular curriculums, must include opportunities for teachers to increase subject matter mastery, must include teachers observing and commenting on each others' teaching, and must be an ongoing part of teachers' work.

Recent evidence from Texas shows that professional development that includes these components can lead to better instruction and improved student achievement (Holcombe 2002). As documented in a well-designed evaluation and shown in Figure 3, participation in a state-sponsored Algebra Institute led to improved student achievement. A related finding demonstrates that the more teachers in a school participated in the training, the greater the impact on students' achieve-





ment. This evidence is important because it documents that professional development can make a difference to student achievement.

#### Suggestion Seven: Show the Way on Summer Learning

Recent evidence documents a pattern that many educators have suspected for a long time—namely, that the cognitive skills of lowincome children fall compared to those of middle-class children over the summer months. This suggests that developing rich summer learning opportunities may be a powerful way to increase the cognitive skills of low-income children. It may make sense for teacher unions to lead the way (perhaps in conjunction with local foundations) in designing summer programs and in funding independent evaluations of their effectiveness.

#### Policy Advice for Improving Urban Schools

While the above comments are aimed at teacher union leaders, they have direct implications for advice to local and state policymakers. In particular, for these policymakers, the seven points listed above can be recast as follows:

1. Retain a single-minded focus on improving student achievement; build constituencies among business groups and community groups supporting this unwavering focus.

- 2. Invest in developing well-defined content standards and in developing student assessments that are tightly aligned with standards.
- 3. Create a level playing field on which public schools compete with charter and voucher schools.
- 4. Provide every school with the resources and flexibility needed to succeed (schools serving difficult-to-educate children need significantly more resources).
- 5. Provide all schools with tools, training, and time to learn from student assessments.
- 6. Create conditions that facilitate improvements in instruction; monitor progress in student achievement closely and intervene when progress is not forthcoming.
- 7. Develop, implement, and evaluate programs to provide summer learning to low-income children.

#### References

- Autor, David H., Frank Levy, and Richard J. Murnane. 2001. "The Skill Content of Recent Technological Change: An Empirical Exploration." NBER Working Paper No. 8337 (June).
- Holcombe, Lee. 2002. "Teacher Professional Development and Student Learning of Algebra: Evidence from Texas." Harvard Graduate School of Education, doctoral thesis (June).U.S. Census Bureau. 1989. *Historical Statistics of the United States: Colonial Times to 1970*. Part
- U.S. Census Bureau. 1989. *Historical Statistics of the United States: Colonial Times to 1970*. Part 1. White Plains, NY: Kraus International Publications.
- U.S. Department of Education, National Center for Education Statistics. 1983. Condition of Education, 1982. Washington, DC.
- 2001. Private School Universe Survey, 1999–2000. NCES 2001–330, by S. P. Broughman and L. A. Colaciello. Washington, DC.
- ——. 2002. Digest of Education Statistics, 2001. NCES 2002–130, by T. D. Snyder. Washington, DC.

## Panel Discussion

## AN EDUCATION SUPPORT SYSTEM

### Warren Simmons\*

Before I begin the substance of my remarks, I would like to reveal some of my background because it will explain the views I am about to share. I am usually described as an expert on urban education. During the past ten years I have been involved in urban school reform, including efforts as a researcher, central office administrator, or leader of a local education fund in Prince George's County, Maryland, Baltimore, Philadelphia, and Washington, DC public schools. All of these systems are now subject to some type of state intervention. Thus, you might not be pleased to learn that I am now developing a close relationship with the Boston public school system. We will see if my record remains consistent.

More seriously, my experience has taught me several lessons about the magnitude and complexity of implementing standards-based reform in urban school systems. But first, I would like to point out that much of the conference discussion has dealt with standards-based accountability—a part, but not the whole, of standards-based reform. Standardsbased reform emphasizes the importance of using content and performance standards—descriptions of what students should know and be able to do along with examples of what it means to be proficient—to strengthen and align curriculum and instruction, assessment, and professional development, as well as to inform decisions about school funding and other factors central to teaching and learning. This movement, which began in the 1990s, was spurred by international comparisons of student performance that showed American students faring poorly in mathematics and science in relation to their peers in Europe and Asia. The validity of these findings has been challenged based on differences in the range of

<sup>\*</sup>Executive Director, Annenberg Institute for School Reform, Brown University.

students tested from country to country. The results can also be tied to variations in the nature of the education system in the United States compared to the ones in place in Europe and Asia. That is, these studies are also examining the effects of national education systems that exist in most European and Asian countries with the results of the federal system present in the United States.

In most national systems academic standards are set at the national level by a government education authority that also has the power to develop assessment, curriculum, and professional development strategies that are aligned with the nation's standards. In the federal system employed in the United States, education is controlled by the states, which in turn, delegate authority to over 16,000 separate school districts. The federal government encourages each state to develop its own standards and assessments. And while most states have complied, curriculum and professional development tend to be designed and implemented through the combined efforts of higher education institutions, textbook publishers, and school districts. Achieving alignment among standards, assessment, curriculum, and instruction in our "loosely coupled" federal system (Elmore 2000) is a more difficult and complex enterprise than in most national systems, given the distribution of roles and responsibilities among federal, state, and local education authorities and providers.

To date, federal and state efforts have paid far more attention to the development of standards-based assessments and accountability systems, compared with the amount of attention devoted to curriculum and instruction. As a result, our nation's ability to measure student progress against a collection of state standards that vary considerably exceeds our ability to provide learning opportunities to give all students the support they need to reach the standards, if they work hard enough. In short, because of the variance caused by state and local differences in standards, assessment, curriculum, and instruction, one should expect a broader distribution of student achievement in countries like ours that have federal systems of education. This leads to an important question: Can our nation produce the uniformly high results being demanded by the latest iteration of standards-based reform—No Child Left Behind—given the kind of system we have in place?

Let's refresh our memories somewhat about standards-based reform. The 1993 version of standards-based reform represented by Goals 2000 maintained that if states or national organizations developed academic standards, embedded them in assessments, and attached consequences to performance, the data and pressure generated would improve instruction by guiding the policies and practices of educators and decision makers at the state and local level. Since 1993, federal policy has been inching states and districts closer to the realization of the accountability portion of standards-based reform, while respecting state and district discretion to choose a range of measures to address teaching and learning.

By the end of the decade—that is, the year 2000—we learned that in most respects we were no closer to reaching our national education goals than we were at the beginning of the journey. Through the No Child Left Behind Act, we are continuing our standards-based reform journey with a more ambitious set of goals and a new timeline—the goals must be met by the year 2014. And the question is: When that date arrives, will we still be short of our goals, without paying any consequences for our failure to meet them? I believe the answer to this question depends on the degree to which we remain more preoccupied with the narrow agenda of standards and accountability, as opposed to the opportunity-to-learn side of standards-based reform.

The broader, richer version of standards-based reform that I think exists in many of the countries to which we compare ourselves looks like this: Standards are used to inform, align, and create greater coherence among curriculum, instruction, and assessment. If a state or a country has standards embedded in its tests, those same standards should be used to guide curriculum and instruction. Moreover, those same standards should inform professional development, that is, how one develops expertise among teachers, principals, and central-office staff. In addition, standards should be used to inform how schools are organized and to advance evidence-based discussions with the public about the kinds of resources students need so that excellence becomes a feasible goal for all students, not just the privileged few.

The disappointing results produced by the accountability portion of standards-based reform have fostered a desire to go beyond measuring what students can do, to building a better understanding of what it will take to get them to a destination of high performance. An increasing number of districts are defining a core set of practices that are needed to give all students a fair shot at meeting the standards or at least performing well on the standardized tests that have become proxies for the standards. While this approach represents a step toward paying more attention to the relationship among standards, assessment and accountability, and opportunity-to-learn, it still leaves us short of employing all the tools necessary to foster the alignment envisioned by standards-based reform. Again, the countries we compare ourselves to, such as the United Kingdom, not only have assessment systems and standards, but also use those assessments and standards to guide professional development, curriculum development, funding, public engagement, and conversations about school organization.

In this country, standards and assessments exist alongside a baffling array of recommended school- and student-improvement strategies, such as using state or district standards to examine student work and assignments, tutoring and mentoring students, implementing contentbased and classroom-embedded coaching for teachers, school reconstitution, providing school choice for students in failing schools, creating small autonomous schools, adopting scientifically based reading and math programs, and implementing research-based comprehensive school designs, to name just a few. When one goes into schools to ask if people are doing most of the things on this list, the answer is usually yes. Further probes, however, quickly reveal that few people understand the relations among the strategies on the above list. Moreover, the strategies tend to be treated as approaches that are adopted in serial fashion—that is, a school will adopt a comprehensive school design for a year or two, and then drop it when the district mandates a research-based reading or mathematics program. Moreover, few people can explain what any of this has to do with a state's or a district's standards.

It seems that people fall into four broad groups. The first group is fixated on an accountability model that they believe will drive change by changing the standards, embedding them in assessments, and providing sanctions and awards. Another group believes that experts embedded in comprehensive school designs can drive improvement and change. A third group argues for charters, privatization, and increasing competition. And we still have, existing alongside all of this, the traditional compliance model, where we just have rules and regulations, which, if properly enforced, will cause change to occur.

I submit that none of these models in and of itself will produce change at a massive scale as required by No Child Left Behind. The rate and levels of improvement required by this Act call for a systematic and focused effort that goes far beyond the cherry-picking of school reform strategies in which most schools and districts engage as a response to pressures to improve. The failure of this approach is underscored in the National Assessment of Educational Progress (NAEP) results over the last 20 years. Despite good intentions and efforts, the achievement gap between white students and their black and Hispanic counterparts remains virtually unchanged. Moreover, the vast majority of all students fail to meet proficient levels of performance as defined by NAEP. Michael Barber's conference presentation showed that proficient levels of performance are being met in the United Kingdom (see Barber 2002). By contrast, the percentage of students in the United States who score proficient or above on NAEP is close to 25 percent of white students, but only 4 or 5 percent of African Americans and 3 percent of Hispanics. Miniscule percentages of minority students perform at the proficient level.

What we have been able to do with the current system is to get large numbers of poorly performing students at the elementary and middle school levels to approach basic levels of performance. But we have made very little progress in changing student performance in our nation's high schools. We have to ask ourselves if we can meet the expected rate of improvement through the federalist education system that we now have, a system in which we essentially ask teachers, students, and parents in individual schools to take on the task of responding to the information produced by state assessments (which they rarely receive in a timely fashion), while districts, higher education institutions, providers, and the states offer a dizzying and sometimes conflicting array of "best practices."

The No Child Left Behind (NCLB) Act continues this trend by suggesting a wide range of improvement strategies, while narrowly defining who should improve and by when. NCLB requires states, districts, and schools to meet annual improvement objectives or face consequences for failure. Moreover, the failure will be public rather than private; failing schools and districts will be identified and a host of now-familiar actions will be taken. Students and teachers can be transferred. New curricula will be adopted. Some schools will be reconstituted. And in some cases, local superintendents and school boards will be dismissed. In other words, we will continue to do the kinds of things that we have been doing for the last decade with little effect.

I think our efforts to improve schools have been hampered by an incomplete definition of the problem. Whether we take a top-down or bottom-up view of school failure, we tend to define the problem as something inherent in an individual school rather than the product of the larger system. Even when we view the system as the problem, the solution tends to be "freeing" individual schools from the system with the hope that this will force the system to change or collapse. I think we need to change our focus substantially. We must begin to think not just about how to build the capacity of individual schools, but also about how we build, redesign, and reconstruct a local education support system—a system that would support a community or portfolio of schools. No matter where one falls on the spectrum of school reform (on issues such as privatization, choice, or change within the current system), meeting the goals of NCLB requires supports that will dramatically improve performance across a community of schools simultaneously and continuously.

My experience leads me to believe that several factors must be addressed for communities to develop the capacity to take on this work. To begin, communities need cross-sector leadership development to align the vision and efforts of local decision-makers. An increasing number of mayors and city-council people are directly involved in this work, joined by school-board members and superintendents. Mayors and city-council members often come to this work with very little background in education reform. We seem to be enamored with hiring people with management expertise outside of education to lead schools (such as generals, former CEOs, and attorneys), but we have neglected to provide ways for them to acquire the information and knowledge they lack, other than by learning on the job. Local cross-sector leadership development is essential to ensure that major decision-makers create a shared understanding of the nature of the problem and its solution. This kind of leadership development must be informed by data and research on local conditions of instruction buttressed by national research. Currently, local leaders are inundated by the latter, but lack the information they need to adapt national models to fit local circumstances and approaches. For instance, organizations like the Boston Plan for Excellence take national designs and models, customize them, and work with local educators, parents, and community members to heighten their ability to support effective implementation.

Effective public engagement is another critical need at the local level. Many school superintendents, teachers, and principals do a poor job of communicating with the public and of engaging members of the public as partners. As several previous speakers mentioned, if we are going to require resources and supports from individuals, groups, agencies, and organizations outside of schools, then we must find ways to communicate with and engage people from a variety of sectors in defining problems and their solutions. As part of this endeavor, we must address ways to change governance to ease the degree to which expertise and resources from cultural institutions, social services, recreation, juvenile justice, child welfare, employment, and education might be pooled and applied to increase supports for learning inside and outside of school.

I think the conversation about system change rather than simply school change is beginning to increase in volume despite our culture's resistance to thinking about education in this way. I urge each of you to join this conversation about how we build a local infrastructure, not a school district necessarily, but a local infrastructure with the capacity to make our national education goals a reality rather than a hollow promise.

#### References

Barber, Michael. 2002. "The Challenge of Transformation," this volume.

Elmore, Richard F. 2000. *Building a New Structure for School Leadership*. Washington, DC: The Albert Shanker Institute.
## About the Authors

DARON ACEMOGLU is Professor of Economics at the Massachusetts Institute of Technology. Before joining the faculty at MIT, he was a lecturer in economics at the London School of Economics. Acemoglu's fields of interest include income and wage inequality, human capital and training, economic growth, technical change, labor economics, search theory, and political economics. The recipient of numerous fellowships and grants, he has published in the *American Economic Review*, the *Journal of Economic Literature*, and other scholarly journals. His most recent article, "Directed Technical Change," is forthcoming in the *Review of Economic Studies*. Acemoglu received both his M.Sc. and his Ph.D. from the London School of Economics.

MICHAEL BARBER is Head of the British Prime Minister's Delivery Unit and Chief Adviser to the Prime Minister on Delivery. His role is to ensure that the government has in place the plans and capacity to deliver its priorities in four areas of public service: health, education, law and order, and transport. Until June 2001, he was head of the Standards and Effectiveness Unit at the Department for Education and Skills and chief adviser to the British Secretary of State for Education and Skills on School Standards. In this role, Barber was responsible for the implementation of the government's school reform agenda. His work has been published widely in both academic journals and the press, and he speaks regularly on radio and television about education policy. After attending Bootham School in York, Barber studied at Oxford University and the Georg August University in Göttingen, Germany.

JULIAN R. BETTS is Professor of Economics at the University of California at San Diego. He is also a senior fellow at the Public Policy Institute of California, where he helps to direct the Institute's K–12 education research effort. Much of his research has focused on the economic analysis of public schools. Betts has written extensively on the link between student outcomes and measures of school spending, including class size, teachers' salaries, and teachers' level of education. More recently, he has examined the role that standards and expectations play in student achievement. His other main areas of research include higher education; immigration; the economics of unions; and technology, skills, and the labor market. In 2001, Betts was one of 14 people appointed to the National Working Commission on Choice in K–12 Education. Betts holds an M.Phil. in economics from Oxford University and a Ph.D. in economics from Queen's University, Ontario.

JOHN H. BISHOP is Associate Professor of Human Resource Studies at Cornell University. Additionally, he is executive director of the Educational Excellence Alliance, a consortium of over 400 high schools that is studying ways to improve student engagement and school climate. He has published numerous articles on education reform, the impact of education quality on individual and national productivity, and the impact of business policies on creating incentives for student achievement. A recent article by Bishop, "Why Do Students Learn More When Achievement Is Examined Externally?" is forthcoming in *EdMatters*. Bishop's current research focuses on the national and international development of curriculum standards and testing to hold school systems accountable. Bishop holds a Ph.D. in education from the University of Michigan.

PETER J. DOLTON is Professor of Economics at the University of Newcastle and at the Institute of Education at the University of London. He is also a research fellow

at the Centre for the Economics of Education at the London School of Economics. He has had visiting fellowships at the University of Paris and the University of Amsterdam. He is currently an associate editor of the *Economics of Education Review* and *Education Economics*. His research interests include the economics of education, labor economics, and applied econometrics. His work has focused on the graduate labor market, teacher supply, discrimination, job tenure, survival data, and the youth labor market. Dolton's current projects involve the teacher labor market, teacher pay and performance, and school inputs and outputs. He has consulted for the Lord Chancellor's Department, the World Bank, and the OECD. Dolton received his Ph.D. from the University of Cambridge.

THOMAS A. DOWNES is Associate Professor of Economics at Tufts University and was a visiting scholar at the Federal Reserve Bank of Boston. His major research interests include the evaluation and construction of state and local policies to more efficiently and equitably deliver publicly provided goods, with particular attention paid to the provision of public education. Downes has researched the roles of the public and private sectors in the provision of education and the potential outcomes of school choice, as well as the impact on student outcomes of various school finance reforms. His published research has appeared in such journals as the *Public Finance Review*, *National Tax Journal*, *RAND Journal of Economics*, and *Journal of Urban Economics*. Downes received his Ph.D. from Stanford University.

DAVID N. FIGLIO is Walter J. Matherly Professor of Economics at the University of Florida and a faculty research fellow at the National Bureau of Economic Research. Prior to joining Florida's faculty, Figlio was an assistant professor at the University of Oregon. Figlio's research focuses on topics in the economics of education and social policy and on the political economy of policy formation and implementation. Much of his current research concerns the unintended consequences of school accountability systems. Additionally, he is studying outcomes as disparate as school lunch programs, student disability classification, school zoning, and house prices. Figlio's research has been published extensively in leading journals, including the *American Economic Review*, the *Journal of Public Economics*, and the *Journal of Law and Economics*, as well as several policy-oriented outlets. He received his Ph.D. in economics from the University of Wisconsin-Madison.

CHESTER E. FINN, JR., is President of the Thomas B. Fordham Foundation and is John M. Olin Fellow at the Manhattan Institute. His primary focus of concern in both these capacities is the reform of primary and secondary schooling. Finn is also a fellow of the International Academy of Education, a distinguished visiting fellow at Stanford's Hoover Institution, and an adjunct fellow at the Hudson Institute. Earlier, Finn was founding partner and senior scholar at the Edison Project, and he served as assistant secretary for research and improvement at the U.S. Department of Education. Finn has written extensively on educational reform; he is the author of over 13 books and 300 articles. Most recently he has co-edited *Rethinking Special Education for a New Century*. Finn received his master's degree in social studies teaching and his doctorate in education policy, both from Harvard University.

CLAUDIA GOLDIN is Henry Lee Professor of Economics at Harvard University and director of the Development of the American Economy program at the National Bureau of Economic Research. Before joining the faculty at Harvard, she taught at Princeton University and the University of Pennsylvania. Her research has been funded by the National Science Foundation and the Spencer Foundation, and she was a Guggenheim Fellow. Goldin's research is in the general area of American economic history and has covered topics that range widely, including slavery, women in the economy, technological change, and most recently, education. She is the author and editor of several books and numerous academic articles. Her most recent work is on the rise of mass education in the United States and its impact on inequality; this work will form the core of her book in progress, *The Race between Education and Technology*. Goldin received her M.A. and Ph.D. in economics from the University of Chicago.

ELLEN C. GUINEY is Executive Director of the Boston Plan for Excellence, an education foundation that supports Boston's public schools in their transformation of core teaching activities, their efforts to make more strategic use of existing resources, and their accountability to parents and the public. She also serves as co-director of the Boston Annenberg Challenge. As executive director of the Boston school district's primary partner in its reform effort, Guiney provides leadership for and oversees the work of 26 "effective practice" schools. Prior to joining Boston Plan for Excellence, Guiney was chief education advisor to the U.S. Senate Committee on Labor and Human Resources. She also served as education advisor to Boston Mayor Raymond L. Flynn during the transition to an appointed school board. Guiney is a former high school English teacher with degrees from Boston College and LeMoyne College.

ERIC A. HANUSHEK is Paul and Jean Hanna Senior Fellow at the Hoover Institution of Stanford University, as well as a research associate at the National Bureau of Economic Research. He is a leading expert on educational policy, specializing in the economics and finance of schools. Hanushek's books include *Improving America's Schools, Making Schools Work*, and *Education and Race*. In addition, he has published numerous articles in professional journals. He previously held academic appointments at the University of Rochester, Yale University, and the U.S. Air Force Academy. His government service includes posts as deputy director of the Congressional Budget Office, senior staff economist for the Council of Economic Advisers, and senior economist for the Cost of Living Council. Hanushek is a distinguished graduate of the United States Air Force Academy and holds a Ph.D. in economics from the Massachusetts Institute of Technology.

ROBERT H. HAVEMAN is John Bascom Professor in the Department of Economics and the LaFollette Institute of Public Affairs at the University of Wisconsin-Madison. He is also a research affiliate at the University's Institute for Research on Poverty, and was appointed adjunct research professor at Australian National University. Haveman served as senior economist on the Subcommittee on Economy in Government, Joint Economic Committee, U.S. Congress. He was a fellow at the Netherlands Institute for Advanced Study and at the Russell Sage Foundation, and research professor at the Brookings Institution. Haveman has published widely in the fields of public finance, benefit-cost analysis, and the economics of poverty and social policy. His work has appeared in numerous professional and academic publications, including the *American Economic Review*, the *Quarterly Journal of Economics*, and *Journal of the American Statistical Association*. Haveman holds a Ph.D. in economics from Vanderbilt University.

THOMAS J. KANE is Professor of Policy Studies and Economics at the University of California at Los Angeles. Much of his work has focused on topics in higher

education—estimating the labor market payoff to a community college education, estimating the impact of tuition and financial aid policy on college enrollment rates, and estimating the impact of affirmative action in college admissions. Kane's book, *The Price of Admission: Rethinking How Americans Pay for College*, examines state and federal financial aid policies. His current work focuses on the design of accountability systems in elementary and secondary education. Prior to his affiliation with UCLA, Kane was an associate professor of public policy at Harvard University. Additionally, he has been a visiting fellow at the Brookings Institution and at the Hoover Institution at Stanford University. Kane served as senior staff economist for labor, education, and welfare policy issues for President Clinton's Council of Economic Advisers. Kane holds a Ph.D. in public policy from Harvard University.

LAWRENCE F. KATZ is Professor of Economics at Harvard University and a research associate at the National Bureau of Economic Research. His research involves topics in the general areas of labor economics and the economics of social problems, including wage and income inequality, unemployment, the economics of education, and the evaluation of the effectiveness of social and labor market policies. Katz is the author of numerous articles in scholarly journals. Currently, he is examining the role of technological change and the pace of educational advance in affecting the wage structure. Katz has been editor of the *Quarterly Journal of Economics* for over ten years. He served as chief economist of the U.S. Department of Labor and was the first director of the program on children at the National Bureau of Economic Research. Katz holds a Ph.D. in economics from the Massachusetts Institute of Technology.

YOLANDA K. KODRZYCKI is Assistant Vice President and Economist at the Federal Reserve Bank of Boston, where she specializes in regional, labor, and public sector economics. Her most recent research has examined migration of recent college graduates and causes of regional differences in educational attainment. Kodrzycki serves as an advisor to the Massachusetts Technology Collaborative, the Massachusetts Division of Employment and Training, and the New England Board of Higher Education, and she is a member of the editorial board of *Massachusetts Benchmarks*, an economics magazine issued by the University of Massachusetts. She is a past president of the New England Economic Project, a nonprofit organization dedicated to analyzing and forecasting the regional economy. Kodrzycki served as consultant to the U.S. Treasury advisory program in Central and Eastern Europe. Prior to joining the Boston Fed, she taught economics at Amherst College. Kodrzycki received her Ph.D. in economics from the University of Pennsylvania.

ALAN G. MERTEN is President of George Mason University. Previously, he was dean and professor of information systems at the Johnson Graduate School of Management at Cornell University. He was also affiliated with the University of Florida and the University of Michigan. Merten serves as a director of the Greater Washington Board of Trade, INOVA Health System, several information technology companies, and a mutual fund trust. He has been recognized for his contributions to the Northern Virginia technology community and for his contributions to the use of information technology in the federal government. Merten received his master's degree in computer science from Stanford University and holds a Ph.D., also in computer science, from the University of Wisconsin. CATHY E. MINEHAN is President and Chief Executive Officer of the Federal Reserve Bank of Boston. As one of the nation's central bankers, she contributes to policy decisions that promote the safety and soundness of the U.S. financial system and the health of the nation's economy. Minehan is an expert in payment systems, a major Fed responsibility, and currently chairs the System's Financial Services Policy Committee, which directs the strategic efforts of the Reserve Banks in providing payments and other services. She also serves on the boards of many civic, professional, and educational organizations, including the Boston Private Industry Council, Jobs for Massachusetts, the United Way, the University of Rochester, and Bentley College. Minehan began her career with the Federal Reserve System following graduation from the University of Rochester. She received her M.B.A. from New York University and holds several honorary degrees.

RICHARD J. MURNANE is Thompson Professor of Education and Society at Harvard University's Graduate School of Education. An economist who specializes in education issues, he has written extensively about how changes in the economy affect the education sector. Murnane's books include *The Impact of School Resources on the Learning of Inner City Children, Who Will Teach: Policies That Matter,* and, with his colleague, Frank Levy of MIT, *Teaching the New Basic Skills.* Levy and Murnane are currently completing a book describing how the computerization of work makes some human skills more important and others less important. Murnane is also currently working half-time with the Boston public schools, helping schools to make more effective use of student assessment data in diagnosing students' achievement problems and planning instructional interventions. Murnane received both his M.A. and his Ph.D. in economics from Yale University.

MARGARET E. RAYMOND is Research Fellow at the Hoover Institution at Stanford University. There, she directs CREDO (formerly the Center for Research on Education Outcomes), which analyzes education reform efforts around the country. The Center originated at the University of Rochester, where Raymond was associate professor of political science and public policy analysis. Raymond has done extensive work in public policy and education reform and is currently researching the development of competitive markets and the creation of reliable data on program performance. Raymond's extensive list of research publications includes the first national evaluation of *Teach for America*. In addition to her work at CREDO, she has served as consultant to federal, state, and local agencies. Raymond holds a Ph.D. in political science from the University of Rochester.

MICHAEL A. REBELL is Executive Director and Counsel of the Campaign for Fiscal Equity, Inc. An experienced litigator, administrator, researcher, and scholar, he specializes in the field of education law. His research includes an ongoing examination of fiscal equity and adequacy reform issues throughout the nation and the use of public engagement techniques to effectively implement institutional reform remedies in class action litigations. Rebell is the co-author of two books, *Equality and Education* and *Educational Policy Making and the Courts*, and has authored dozens of articles on issues of law and education. Rebell is a frequent lecturer and consultant on education law. He has taught courses on law and education for many years at the Yale Law School, and currently he is an adjunct professor and lecturer in law at Teachers' College and Columbia Law School. Rebell holds an L.L.B. from Yale Law School.

T. PAUL SCHULTZ is Malcolm K. Brachman Professor of Economics at Yale University. He is affiliated with the Economic Growth Center at Yale, and previously served as its director. Prior to joining the faculty at Yale, Schultz was a professor of economics at the University of Minnesota. His current research interests include schooling, health, and mobility in development; income distribution and endogenous household composition; and gender inequalities. He has written numerous articles for professional and academic publications on these subjects. One of his recent articles, "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program," is forthcoming in the *Journal of Development Economics*. Schultz received his Ph.D. from the Massachusetts Institute of Technology.

WARREN SIMMONS is Executive Director of the Annenberg Institute for School Reform at Brown University. The Institute strives to improve conditions and outcomes in American schools, particularly in urban areas and in schools serving disadvantaged students. Before joining the Institute, Simmons was executive director of the Philadelphia Education Fund, a nonprofit organization that helped the Philadelphia school district fund, develop, and implement new academic standards, content-based professional development, standards-based curriculum resources, and comprehensive school reform, as part of the Children Achieving reform agenda. Simmons serves on the boards and advisory groups of numerous education-reform initiatives, including the Public Education Network and the Cross-City Campaign for Urban School Reform. His publications have addressed the interaction of culture, cognitive development, and achievement, as well as literacy technologies and assessment. Simmons received his Ph.D. in psychology from Cornell University.

PAULO RENATO SOUZA is Minister of Education for the federal government of Brazil. Previously, he was operations manager for Inter-American Development Bank in Washington, D.C., president of the State University of Campinas, and secretary of education for the state government of São Paolo. As professor of economics, Souza has taught at several prestigious institutions, including State University of Campinas, Catholic University of São Paolo, and the Federal University of Rio de Janerio. He was the invited director of studies at Ecole de Hautes Etudes en Sciences Sociales in Paris and a visiting researcher at the Institute of Advanced Study in Princeton, New Jersey. He has written various articles on education, economic development, and labor economics. Souza earned a doctorate in economics from the State University of Campinas.

BARBARA L. WOLFE is Professor in the Departments of Economics and Population Health Sciences and for the LaFollette School of Public Affairs at the University of Wisconsin–Madison. She is also a research associate at the National Bureau of Economic Research. Since joining the faculty at Wisconsin, Wolfe conducted research abroad as a visiting professor at the University of Amsterdam, a visiting fellow at Australian National University, and a fellow-inresidence at Netherlands Institute for Advanced Study. Her areas of interest include health economics and public economics. She has written widely on these topics and is the editor and author of several books and numerous articles in professional and academic journals, including the *American Economic Review*, the *Journal of Public Economics*, and the *International Journal of Educational Research*. Wolfe received both her M.A. and her Ph.D. in economics from the University of Pennsylvania.

## **Conference Participants**

K. DARON ACEMOGLU, Massachusetts Institute of Technology DEBORAH ALLINSON, Wellington Management Company MICHAEL BARBER, Prime Minister's Delivery Unit, United Kingdom JULIAN R. BETTS, University of California, San Diego JOHN H. BISHOP, Cornell University KATHARINE L. BRADBURY, Federal Reserve Bank of Boston FREDERICK S. BREIMYER, State Street Corporation LYNN E. BROWNE, Federal Reserve Bank of Boston VITORIA ALICE CLEAVER, Ministry of Education, Brazil PAUL M. CONNOLLY, Federal Reserve Bank of Boston E. GERALD CORRIGAN, Goldman, Sachs & Company J. DEWEY DAANE, Vanderbilt University PETER J. DOLTON, University of Newcastle on Tyne THOMAS A. DOWNES, Tufts University **ROBERT DUGGER, Tudor Investment Corporation** DONNA DULSKI, Federal Reserve Bank of Boston WILL EDWARDS, Bloomberg News DAVID N. FIGLIO, University of Florida CHESTER E. FINN, JR., Thomas B. Fordham Foundation MARY C. FITZGERALD, Federal Reserve Bank of Boston PATRICIA M. FLYNN, Bentley College JEFFREY C. FUHRER, Federal Reserve Bank of Boston CLAUDIA GOLDIN, Harvard University ALLAN E. GOODMAN, Institute of International Education JEROME H. GROSSMAN, M. D., Harvard University WILLIAM GUENTHER, Mass Insight SCOTT GUILD, Federal Reserve Bank of Boston ELLEN GUINEY, Boston Plan for Excellence TOYOO GYOHTEN, Institute for International Monetary Affairs JANE HANNAWAY, The Urban Institute ERIC A. HANUSHEK, Hoover Institution, Stanford University ROBERT H. HAVEMAN, University of Wisconsin THOMAS M. HOENIG, Federal Reserve Bank of Kansas City THOMAS J. KANE, University of California at Los Angeles JANE KATZ, Federal Reserve Bank of Boston LAWRENCE F. KATZ, Harvard University YOLANDA K. KODRZYCKI, Federal Reserve Bank of Boston CAROL KOLENIK, The Harvard Bridge to Learning and Literacy Program THOMAS L. LAVELLE, Federal Reserve Bank of Boston KRIS LOCKE, The Harvard Bridge to Learning and Literacy Program ANDRÉ MAYER, Associated Industries of Massachusetts ALAN G. MERTEN, George Mason University CATHY E. MINEHAN, Federal Reserve Bank of Boston RICHARD J. MURNANE, Harvard University JAMES J. NORTON, AFL-CIO DIMITRI B. PAPADIMITRIOU, Bard College SCOTT E. PARDEE, Middlebury College

NICHOLAS S. PERNA, Perna Associates WILLIAM POOLE, Federal Reserve Bank of St. Louis MICHAEL J. PRELL, Arlington, Virginia RALPH RAGSDALE, Federal Reserve Bank of Boston MARGARET E. RAYMOND, Hoover Institution, Stanford University MICHAEL A. REBELL, ESQ., Campaign for Fiscal Equity, Inc. SUSAN E. RODBURG, Federal Reserve Bank of Boston T. PAUL SCHULTZ, Yale University SAM SHUMAN, Federal Reserve Bank of Boston WARREN SIMMONS, Brown University ALLEN SINAI, Decision Economics, Inc. PAULO RENATO SOUZA, Ministry of Education, Brazil DAVID STEINER, Boston University WILLIAM O. TAYLOR. The Boston Globe VICTORIA THIEBERGER, Reuters **ROBERT TRIEST, Federal Reserve Bank of Boston** KIMBERLY UNDERHILL, Federal Reserve Bank of Boston RICHARD C. WALKER III, Federal Reserve Bank of Boston RICHARD C. WHITE, Community National Bank ALBERT M. WOJNILOWER, Monitor-Clipper Partners BARBARA L. WOLFE, University of Wisconsin PAUL WONNACOTT, Potomac, Maryland DAVID Wyss, Standard & Poor's

JOYCE ZICKLER, Board of Governors of the Federal Reserve System

THE FEDERAL RESERVE BANK OF BOSTON CONFERENCE SERIES

No. 1	Controlling Monetary Aggregates	June 1969
No. 2	The International Adjustment Mechanism	October 1969
No. 3	Financing State and Local Governments in the Seventies (out of print)	June 1970
No. 4	Housing and Monetary Policy (out of print)	October 1970
No. 5	Consumer Spending and Monetary Policy: The Linkages (out of print)	June 1971
No. 6	Canadian–United States Financial Relationships (out of print)	September 1971
No. 7	Financing Public Schools (out of print)	January 1972
No. 8	Policies for a More Competitive Financial System	June 1972
No. 9	Controlling Monetary Aggregates II: The Implementation	September 1972
No. 10	Issues in Federal Debt Management	June 1973
No. 11	Credit Allocation Techniques and Monetary Policy	September 1973
No. 12	International Aspects of Stabilization Policies (out of print)	June 1974
No. 13	The Economics of a National Electronic Funds Transfer System	October 1974
No. 14	New Mortgage Designs for Stable Housing in an Inflationary Environment	January 1975
No. 15	New England and the Energy Crisis (out of print)	October 1975
No. 16	Funding Pensions: Issues and Implications for Financial Markets	October 1976
No. 17	Minority Business Development	November 1976
No. 18	Key Issues in International Banking	October 1977
No. 19	After the Phillips Curve: Persistence of High Inflation and High Unemployment	June 1978
No. 20	Managed Exchange-Rate Flexibility: The Recent Experience	October 1978
No. 21	The Regulation of Financial Institutions	October 1979
No. 22	The Decline in Productivity Growth	June 1980
No. 23	Controlling Monetary Aggregates III	October 1980
No. 24	The Future of the Thrift Industry	October 1981
No. 25	Saving and Government Policy	October 1982
No. 26	The Political Economy of Monetary Policy: National and International Aspects	July 1983
No. 27	The Economics of Large Government Deficits	October 1983
No. 28	The International Monetary System: Forty Years After Bretton Woods	May 1984
No. 29	Economic Consequences of Tax Simplification	October 1985
No. 30	Lessons from the Income Maintenance Experiments	September 1986

No. 31	The Merger Boom	October 1987
No. 32	International Payments Imbalances in the 1980s	October 1988
No. 33	Are the Distinctions between Equity and Debt Disappearing?	October 1989
No. 34	Is There a Shortfall in Public Capital Investment?	June 1990
No. 35	The Financial Condition and Regulation of Insurance Companies	June 1991
No. 36	Real Estate and the Credit Crunch	September 1992
No. 37	Safeguarding the Banking System in an Environment of Financial Cycles	November 1993
No. 38	Goals, Guidelines, and Constraints Facing Monetary Policymakers	June 1994
No. 39	Is Bank Lending Important for the Transmission of Monetary Policy?	June 1995
No. 40	Technology and Growth	June 1996
No. 41	Social Security Reform: Links to Saving, Investment, and Growth	June 1997
No. 42	Beyond Shocks: What Causes Business Cycles?	June 1998
No. 43	Rethinking the International Monetary System	June 1999
No. 44	Building an Infrastructure for Financial Stability	June 2000
No. 45	The Evolution of Monetary Policy and the Federal Reserve System Over the Past Thiry Years: A Conference in Honor of Frank E. Morris	October 2000
No. 46	Seismic Shifts: The Economic Impact of Demographic Change	June 2001
No. 47	Education in the 21st Century: Meeting the Challenges of a Changing World	June 2002

Copies of individual volumes in the conference series may be obtained without charge by writing to the Research Library—D, Federal Reserve Bank of Boston, P.O. Box 2076, Boston, MA 02106-2076. The fax number is (617) 973-4221, and the e-mail address is boston.library@bos.frb.org. A \$10.00 payment (check drawn on a branch of a U.S. bank) is required for 10 or more volumes or 10 or more copies of the same volume.

Beginning with Volume No. 41, conference series are available as full-text PDF files, downloadable from the Bank's public web site: www.bos.frb.org/economics/ conf/index.htm.

Materials may be reprinted from the conference volumes if the source is credited in full, unless it is otherwise noted at the beginning of a paper or discussion. Please send information about any reprinting of materials to Ann Eggleston, Managing Editor, Federal Reserve Bank of Boston.



Federal Reserve Bank of Boston