

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.

*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 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, one-

fifth 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 comput-

ing 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 conver-

gence 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 bottom-performing 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 two-thirds of a grade-level equivalent. He measures performance according to scores on widely applied standardized tests that are not curriculum-based 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 market-driven 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 cost-cutting, 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 ever-increasing 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.