

Should America Be More like Them? Cross-National High School Achievement and U.S. Policy

Author(s): David P. Baker, Jaekyung Lee and Stephen P. Heyneman

Source: *Brookings Papers on Education Policy*, No. 6 (2003), pp. 309-338

Published by: Brookings Institution Press

Stable URL: <http://www.jstor.org/stable/20067259>

Accessed: 11-11-2017 20:40 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



Brookings Institution Press is collaborating with JSTOR to digitize, preserve and extend access to *Brookings Papers on Education Policy*

*Should America Be
More Like Them?
Cross-National High School
Achievement and U.S. Policy*

DAVID P. BAKER

The modern comprehensive American high school, since its inception in the early twentieth century, has been considered alternately an organizational blessing and a bane on educational progress. Welcomed as an organizational advance through which the hodgepodge of schools in urban America could be made into an orderly pedagogical and administrative pyramid, the early modern high school was seen as an educational institution meeting the social and economic challenges of an increasingly diverse industrial-urban society.¹ The image of a rational, bureaucratic, large, and robust comprehensive high school was pushed forward through the middle of the twentieth century as a progressive and modernizing model for the entire nation.² However, by the second half of the century, as evidence of social decline, persistent poverty, racial disparities, and educational failure in urban communities became ever more obvious, the image of the urban comprehensive high school shifted from an exemplary model to a broken institution in need of reform.

The world's image of the American comprehensive high school has run a similar course over the twentieth century. The creation of a comprehensive and socially integrated secondary education was the cornerstone of U.S.

I wish to thank Diane Ravitch for organizing a stimulating policy conference; Stephen Heyneman and Jaekyung Lee for their insightful comments on the paper during the conference; and Brian Goesling and Maryellen Schaub for their assistance on earlier drafts.

education plans for reforming the defeated fascist regimes in Japan and Germany. It also had a significant influence on the development of modern secondary education systems throughout the world from the 1940s on.³ Yet by the time the influential *A Nation at Risk* report was issued some forty years later, secondary education in the most politically and economically powerful nation in the world was declared internationally mediocre at best.⁴ Ironically, many of the organizational features of the once-lauded American high school became the focus of intense speculation about how they might impede educational quality.

At the same time, international comparisons of American education have become popular and widespread in the reform debates of the last several decades. Many centerpieces of the current American reform movement were shaped through evidence and speculation on what other more educationally successful nations were doing that the United States was not. The growing volume of international studies and data available with which to compare the United States has influenced this policy trend. For example, the international data portfolio of the National Center on Education Statistics in the Department of Education has more than tripled since the 1980s and is slated to grow even more over the next decade.

As a consequence, many debates about what should be done to reform the American high school revolve around assessing U.S. national competitiveness in the world's educational progress. This is evidenced by both the Bush administration's early 1990s educational strategy and the Clinton administration's Goals 2000 (see Goal #5) and informally through two decades' worth of op-ed pieces on the international state of American education. So the question has become not only what would improve American secondary education, but also what is required to make it the best in the world? Making international competitiveness a main goal of educational reform has opened the door to wide speculation about what should be imported from other nations to improve secondary education. In other words, in American education policy circles an often-asked question over the past two decades has been: What is it that other nations do to make their mathematics and science education more effective than the United States and should it become more like theirs? Many suggestions have been put forward, and these continue to circulate within American policy debates as possible reasons that some nations do better than others in mathematics and science.

Specific features from other education systems that have intrigued American reformers of secondary education fit into five categories.

1. Create nationwide school and classroom climates that are conducive to teaching and learning.
2. Manage the nation's schools with enough local autonomy for accountability without too many centralized restrictions.
3. Motivate a nation's student body to study more and work harder to learn mathematics and science.
4. Develop and implement world-class mathematics and science curricula for a nation's secondary schools.
5. Train and motivate a national faculty in secondary mathematics and science to teach effectively, and provide them with the necessary school-based resources to do so.

Ironically, given their origin in international comparisons, many ideas for national reform have not undergone any serious cross-national testing of their ability to change a nation's standing in secondary education effectiveness.⁵ Underlying the tendency to make national policy based on untested observations from cross-national data is the belief among the policy community that on the more microlevels, such as schools and students, the factors represented in the five reform categories are strongly related to a student's achievement, and empirical evidence supports this. The problem occurs with the leap to inferring that these factors will also answer the larger question of how the United States could change its overall level of international achievement competitiveness. A subtle, but problematic, shift in causal reasoning takes place when reform analysts move from speculation about what needs to happen at levels below the nation such as schools, classrooms, and students to what needs to happen at the national level. For example, no one would doubt that a demanding curriculum taught by an effective teacher in a school with resources leads to better achievement; or that a motivated student, other things equal, will learn more than an unmotivated one. Hence the assumption is that nations that do these things will have higher average achievement in subjects such as mathematics and science. But the question remains, How can nations make policy to do these things across large numbers of schools? This is often not known for sure, and further, as I show here, often national differences in aggregations of microfactors do not answer the question of what might cause cross-national differences in secondary education in mathematics and science.

Similarly, the average American high school is often assumed to be lacking in central aspects related to achievement in comparison to secondary education in other nations. Inferring backward from the United States' gen-

erally mediocre international performance, the assumption is that many American high schools are ineffective because of clear differences from high schools in more effective nations. Again, for most of the five reform speculations, this is rarely statistically tested and this unconfirmed assumption continues to have considerable impact on policy about how to move U.S. secondary education up in international rankings.

Finally, in a similar vein, in observing the international standing of the American high school, policymakers often note the diversity of students and schools within the large, localized U.S. education system. They assumed that this national feature causes special problems for certain populations of disadvantaged students, but again this has not been empirically examined cross-nationally.

Using the detailed, international data from the original 1994 Third International Mathematics and Science Study (TIMSS) and its replication in 1999, I explore these issues for a selected set of indicators for each of the five reform dimensions.⁶ Although I focus on mathematics, much of what is described here could be applied to science as well. The analyses presented are based on a set of technical articles and papers in which indicators of these dimensions are examined and modeled in multivariate detail.⁷ My colleagues and I have prepared these statistical analyses as part of a U.S. Department of Education–National Science Foundation jointly funded project providing theoretically based and policy-relevant secondary analysis of the 1994 TIMSS data set.

American High School from an International Perspective

Two questions are considered for each of the five secondary education reform issues: How much of a difference exists between the average American high school and the average across nations? Is the difference associated with cross-national achievement? The answers are summarized in table 1.

Reform Issue 1: Create nationwide school and classroom climates that are conducive to teaching and learning.

Many factors are conducive to a learning climate in schools and classrooms. Speculation on the state of American schools and classrooms focuses on an image of frequent disruptions by students resulting in a less effective learning climate. This includes the extreme image of the American high school as a violent and dangerous place. School disruption and vio-

Table 1. American High School from an International Perspective

<i>School characteristic</i>	<i>How much of a difference exists between the average American high school and the average across nations?^a</i>	<i>Is the difference associated with cross-national achievement?</i>
School size	No	No
Math class size	No	No
Disruptive learning environment		
Eighth grade	No	Yes ^b
Twelfth grade	Small	No ^c
Degree of decentralized control of instruction and curriculum (eighth grade)	Large—more decentralized	No
Complexity of decision environment	Small—more complex	Yes—negative
Amount of instructional leadership	Large—less leadership	No
Daily assigned homework	Large—more homework	Yes—negative
Use of remedial shadow education	Large—less use	No—eighth grade Yes—negative, twelfth grade
Classroom implementation curriculum (eighth grade)		
Number of topics covered	Small—more coverage	No
Repetition of topics	Small—more repetitious	No
Instructional resources (eighth grade)		
Absolute amount	No	Yes
Equity of distribution	No	Yes

Source: Third International Mathematics and Science Study, twelfth-grade general mathematics and eighth-grade data.
 Note: Means, correlations, and other statistics and technical details of statistical tests are available from the author upon request.

a. Twelfth-grade nations are Australia, Austria, Canada, Cyprus, Czech Republic, Denmark, France, Germany, Hungary, Iceland, Italy, Lithuania, Netherlands, New Zealand, Norway, Russian Federation, Slovenia, South Africa, Sweden, Switzerland, and United States.

b. Eighth-grade effect vanishes after controlling for national funding level for education and resource inequality.

c. After extreme outlier, South Africa is removed.

lence have been topics of national concern throughout the last decade as shooting incidents have taken place in U.S. public schools. The impact of the media coverage of these occurrences is hard to measure, but reports of student and teacher fears of violence indicate that anxiety is widespread.⁸ Even though some statistical reports show that levels of violence in U.S. schools have generally fallen since the early 1990s, concerns about school violence have increased, not only in the United States but in other nations as well.⁹ Americans tend to perceive their public schools as sites of frequent violence and disruptions, and they probably believe that school violence

occurs more here than in other nations. But how true is this and what effect do levels of student disruptions have on learning cross-nationally?

In an extensive analysis of these issues, Motoko Akiba, Gerald K. LeTendre, Brian Goesling, and I find that American eighth-grade classrooms typically are not disrupted by bullying and threats any more than the average across the other nations in TIMSS.¹⁰ A substantial proportion of American eighth-grade teachers (40 percent) report regular classroom disruptions by students, but this proportion is still below the international average. Although considerable disruptive behavior occurs in the average American middle school, close to the same is occurring on average across other nations. Further, the number of students who report being bullied drops off by the twelfth grade in the United States. In the average American high school, 11 percent of students report being bullied, which is less than the international average of 15 percent. Furthermore, a similar drop-off from middle grades to high school happens in most other nations in the TIMSS study.

In terms of the cross-national relationship between levels of disruption and achievement, although my colleagues and I report a moderate bivariate association of .4 at the eighth grade, the analysis goes on to show that this relationship vanishes when national levels of educational resource equity and funding of education are taken into account. Not even an initial bivariate relationship exists between national achievement and national levels of student disruption in the twelfth grade. Also we find that levels of school disruption and everyday violent student behavior are not related to the level of adult violence across nations. Although, among wealthy nations, the United States has a very high adult violence rate, this does not translate into above average school disruption and everyday violence.

Reform Issue 2: Manage the nation's schools with enough local autonomy for accountability without too many centralized restrictions.

Effective management, local accountability, and lower centralized obstructions are ideas at the heart of much of the recent reform of mathematics and science education in the United States. The notion that, even in the midst of a localized administrative system, many American high schools have too many centralized restrictions (from districts, states, and the federal government) is a popular theme behind the movement for school-based management. The logic behind greater accountability and autonomy to meet curricular and instructional standards and goals is: (1) an effective degree of autonomy at the school level to make curricular and instructional decisions,

(2) a low degree to which other levels of the administrative system can influence (interfere with or complicate) these decisions, and (3) the degree to which the instructional process is under the management of schools. The origins of these prescriptions for more effective schooling are part of a long American policy debate about improving public services such as schools, and images from cross-national studies of schooling are intermixed into this discussion. An image has been created that, compared with schools in other nations, the American school is greatly hampered by ineffective management of curricular and instructional issues.

My colleagues and I have developed three indicators of national management of mathematics curriculum and instruction in schools. These suggest that the international image of the troubled administration of American middle and high schools is perhaps too simple. For example, M. Fernanda Astiz, Alexander Wiseman, and I, using a set of responses from principals, constructed a national measure of the degree to which principals versus central authorities have operational control over decisions and policies about mathematics and science curricula (degree of decentralized control).¹¹ The United States has the least centralized managerial process among all TIMSS nations. This is not just official policy, but also operational practice. Many American principals report that they have considerable control over curricular decisions. At the same time, however, a second measure of managerial complexity for schools shows that the American principal makes curricular and instructional decisions in a more complicated environment than is usual in other nations. In other words, although the American principal tends to have considerable autonomy compared with his or her counterpart in other nations, a number of other American stakeholders (both community and educational authorities) have some degree of influence on school decisions. But cross-nationally this is not as unusual as it might first sound. A widespread global trend in national educational systems also is found toward mixing centralized and decentralized administrative features together. This creates complexity for policymaking in many nations. In this respect the world of education governance may be moving more toward the United States. Wiseman constructed a third managerial measure and finds that the more complex managerial environment in the United States tends to decrease the amount of time and effort principals spend on direct management of instruction in their schools compared with the international average.¹²

The degree to which the indicators of national management of curriculum and instruction are related to national mathematics achievement further

complicates the story. Cross-national variation in decentralization of administration is unrelated to national achievement. But, as Astiz, Wiseman, and I report, more centralized authority does lead to a smaller and more consistently implemented mathematics curriculum across nations' classrooms. Yet neither variation in centralization nor curricular consistency is related to cross-nation variation in achievement. Similarly, although the average American high school principal spends comparatively less time on management of instruction, cross-national variation in the level of instructional leadership is not related to national achievement. Furthermore, a relationship exists between national levels of stakeholders and national achievement, so that nations such as the United States, which tend to have more sources of influence on school decisions, have lower mathematics achievement. These results cast doubt on the notion that decentralization is always the best way to create more effective school management. Optimal international models for school management and achievement are not obvious at this point.

Reform Issue 3: Motivate a nation's student body to study more and work harder to learn mathematics and science.

Two features identified from other nations have received the lion's share of attention about ways to improve American high school students' motivation to study more and work harder to master mathematics at a higher level: (1) high-stakes testing and (2) more academic work outside school, through assigned homework and use of more formal tutoring and other methods known as shadow education. TIMSS affords three useful indicators of motivation and outside schoolwork at the national level.

First is homework. Nationally just over half of American high school students in mathematics have assigned homework every day, which is substantially larger than the international average of just under a third of students with daily homework. While this may be good news for proponents of frequent homework, nations with higher proportions of students assigned homework tend to have lower national mathematics achievement. Gerald LeTendre, Motoko Akiba, Alexander Wiseman, and I find that more homework is often assigned in classrooms to make up for deficient progress in class.¹³ Too much homework, or misuse of homework in many classrooms, may be an indicator of ineffective teaching and a weak application of a curriculum within a nation.

In recent decades the world has seen growing use of tutoring services and related supplemental practices to help students learn the in-school curricu-

lum. This institutional growth has been named shadow education because what goes on in these private services shadows formal school curricula and requirements. Although some nations, such as Japan and South Korea, have elaborate and highly publicized shadow education activities, shadow education was reported in all TIMSS nations. Motoko Akiba, Gerald LeTendre, Alexander Wiseman, and I find that shadow education is widespread, and sizable numbers of students in most nations use it to help them with mathematics.¹⁴ Contrary to earlier speculation about shadow education, most students in most nations use it for remedial purposes (or to address deficits in instruction), not for enhancement. American students' use of shadow education follows this trend and usually is remedial. Furthermore, the overall use of such educational services in the United States is lower than the international average.

Some American educational reformers speculate that the widespread use of shadow education in other nations may lead to higher national achievement—for example, in Japan and South Korea. The main argument is that the presence of extensive shadow education will be positively associated with national achievement levels. The logic behind this idea is twofold. First, extensive shadow education represents extended academic training for a large proportion of the student population and hence has the potential to increase overall achievement. Second, shadow education is a consequence of a larger process that motivates intensive efforts in behalf of certain subjects (for example, mathematics) through such devices as preparing for high-stakes testing or other tightly linked accountability mechanisms.¹⁵ A number of other nations in the mid-1990s, when the TIMSS data were collected, were using high-stakes testing. Although the United States has never had a national high-stakes test of the type found in these nations, since then high-stakes testing has grown as a local and state practice in the United States with the hope that it will increase student motivation to learn.

Internationally, however, not much support is found for this twofold argument as a means to increase national achievement. For eighth grade, my colleagues and I estimated a structural model of the direct effects of each of six detailed indicators of shadow education use in a nation on national mathematics achievement levels plus the indirect effect of high-stakes tests through shadow education. None of the six indicators of shadow education is associated with national variation in mathematics achievement, and the package of high-stakes tests and shadow education is not associated with

variation in national achievement. The analysis also found no direct effect of high-stakes tests on national achievement.¹⁶ At the high school level, wide use of shadow education (again, mostly remedial) is negatively associated with national achievement. These findings indicate that the speculation on national arrangements of high-stakes tests and wide use of shadow education to increase student effort and motivation may not be a cause of cross-national differences in achievement.

Reform Issue 4: Develop and implement world-class mathematics and science curricula for a nation's secondary schools.

A chief messages to emerge from the early reporting on TIMSS is that the United States should reform its mathematics and science curricula if it is to improve internationally. William H. Schmidt and his colleagues have published numerous reports that argue that the intended curricula (that is, the planned official curricula) in U.S. schools are too broad, repetitious, and not demanding enough for American students to compete effectively on a global scale.¹⁷

While I have already examined other factors related to curriculum, such as how decisions about it are made, two indicators of the implementation of the curriculum in eighth-grade general mathematics classrooms are of note. First, Gerald LeTendre, Alexander Wiseman, Erling Boe, Brian Goesling, and I find that the implemented mathematics curriculum in the average American classroom is slightly broader than the average in other nations, but this characteristic of curricula is not related to cross-national achievement.¹⁸ Second, the same is true with the repetition of curricular topics. In the average American classroom there is slightly more repetition, but this characteristic is not related to cross-national achievement.

Reform Issue 5: Train and motivate a national faculty in secondary mathematics and science to teach effectively, and provide them with the necessary school-based resources to do so.

Although the TIMSS data have little to say about the training of mathematics and science teachers, they do contain a significant amount of information about the availability of instructional resources. These data provide information about school-level availability of resources such as budgets for teaching materials, in-classroom supplies, library service and quality, physical plant resources, adequate instructional space, and computer hardware and software. In addition to absolute levels of instructional resources, Brian Goesling, Gerald K. LeTendre, Inga Dora Sigfusdottir, and I developed a measure of instructional resource inequality between schools with which

one can compare them cross-nationally.¹⁹ In other words, a national system of education has large inequality of resources if the distribution across schools is highly uneven; for example, schools serving wealthier communities receive more resources than other schools. Taken together, the absolute and relative resource levels in nations are revealing.

Instructional resources available to the average American school are about the same as the average across the TIMSS nations.²⁰ Similarly, instructional resource inequality in the United States is about the same as in the average TIMSS nation. Both the absolute national level of instructional resources and the nature of their distribution throughout systems are significantly related to national achievement. Further, among the wealthy nations in the TIMSS study, the United States has comparatively high levels of resource inequalities in instructional resources across schools.

Educating Disadvantaged Students in America

A constant challenge facing American education is meeting the special education needs of an economically and culturally diverse society. A number of large, national studies over the past two decades has shown the growth of an American underclass in both urban and rural communities whose conditions are resistant to amelioration by general economic development in the United States.²¹ While the so-called new economy has propelled wealth and opportunity for many, social and economic inequalities are widening. For example, even though there has been a worldwide reduction in income inequality between nations, there is a greater world pattern of within-nation inequality.²²

Numerous critics of American schooling have pointed out the inferior educational opportunities for many disadvantaged youth living in urban and rural communities as well as the significant role limited educational opportunity plays in reproducing poverty.²³ Although, as a result, policy has focused on the poor quality of educational opportunities in public high schools serving disadvantaged students, all the educational improvements and reforms of the last forty years have had less impact on disadvantaged groups in the American society than on the mainstream. Race and social class gaps in achievement continue and remain entrenched within the system.²⁴

With the TIMSS data I examined how well American schools educate students from several types of disadvantaged backgrounds compared with

other nations. I selected two indicators of social and economic disadvantage—low education of the student's mother and living in a single-parent home. Although these family characteristics do not assure a disadvantaged home in every case, they often lead to a disadvantaged home life for many students. The proportion of students with uneducated mothers and single-parent homes in the U.S. twelfth-grade TIMSS sample is 11 percent and 15 percent, respectively (international means for wealthy nations are 29 percent and 11 percent, respectively).²⁵ Among the thirteen wealthy developed nations in the TIMSS twelfth-grade sample, American disadvantaged students learn considerably less general mathematics than similar disadvantaged students in other nations (see table 2). The American mean mathematics knowledge for both students with mothers with less than a high school education and from single-parent homes is strikingly low. Because the U.S. total sample of twelfth graders did not perform well, it is not a surprise that disadvantaged students in the United States also performed poorly. But it is also true that while the American total eighth-grade sample did as well as a number of other wealthy nations, American eighth-grade disadvantaged students finished last compared with similar students in other wealthy nations.²⁶

To illustrate the ramifications of this poor educational record with needy students for national academic competitiveness, one can calculate the impact on the total American mean mathematics knowledge if the United States were to educate its disadvantaged students as well as other nations do. If the United States were as successful as Sweden in educating youth with mothers without a high school degree, the total mean would be improved by 14 points. If the United States did that and also did as well as Sweden does with youth whose mothers have just a high school degree, the U.S. mean would move to 488, which would put the United States beyond five wealthy nations that it currently performs below.²⁷ This is a simple exercise that can be played with any proportion of a distribution. But the message here is clear: Receiving a subpar education significantly adds to the risk of entering the ranks of the underclass, and being from a disadvantaged home in the United States places a youth at greater risk of poor educational performance than in many other nations. Moreover, these findings show that educational systems in other nations have ways to lessen the impact of disadvantaged backgrounds on school achievement. This is not to say that what is done in these nations is perfect or that no negative impact accrues from disadvantaged families. But these findings suggest that social

Table 2. Education of Disadvantaged Students in the United States from an International Perspective

<i>Ranking of nation</i>		<i>Ranking of nation</i>
<i>Students with mothers without a high school degree (mean proportion in nation = 29 percent)</i>	<i>Mathematics mean</i>	<i>Students from single-parent homes (mean proportion in nation = 11 percent)</i>
	560	Netherlands
	550	
Netherlands		Iceland
Sweden	540	Sweden
Denmark		
	530	Switzerland
Iceland		Australia
New Zealand	520	Norway, Denmark
Austria, Norway		Austria, Canada
Switzerland	510	
	International mean	New Zealand
France, Australia	500	
Canada		
	490	
	480	
	470	Germany
	460	
Germany	450	
	440	
United States		United States
	430	

Source: Third International Mathematics and Science Study, twelfth-grade general mathematics data.

and educational policies may be at work in other nations that could help in the United States in meeting the educational challenge of students from disadvantaged families.

What does an international perspective reveal about the policy areas that might be useful to examine in more detail with the aim of improving education for the most disadvantaged? I selected four national educational characteristics from the analyses of my colleagues' and my prior work that had a positive or negative association with cross-national mathematics achievement (see table 3). For each of these characteristics I calculated the national

Table 3. Comparison of U.S. High Schools Serving Students with Mothers without High School Degrees and Similar Schools in Other Wealthy Nations

<i>Characteristic</i>	<i>Comparison</i>
Instructional resources	U.S. schools have more shortages
Equitable distribution of resources	U.S. schools experience more resource inequalities
Daily assigned homework	U.S. schools assign substantially more
Complexity for local decisions about instruction and curriculum	U.S. schools have more stakeholders with influence on decisions

Source: Third International Mathematics and Science Study, twelfth-grade data.
 Note: See table 2 for list of comparison nations.

mean for high schools in the United States and twelve other wealthy nations that enroll the most disadvantaged students (students with mothers who have less than a high school education).²⁸

In each case, the schools that serve American disadvantaged students appear problematic in comparison to high schools in other nations. Schools that serve disadvantaged students in the United States tend to have lower absolute instructional resources and work in a system with greater resource inequalities overall as compared with schools for disadvantaged students in other nations. Given that the level of overall national wealth is similar among these nations, this is not just a function of national differences in wealth. For a wealthy nation, American high schools with disadvantaged students are far behind in supplying fundamental instructional resources.

Teachers in American high schools with disadvantaged students are almost three times more likely to assign daily homework than their counterparts working with disadvantaged students in other wealthy nations. Large amounts of assigned homework could indicate ineffective teaching and curriculum implementation. American principals of these schools are faced with a more complex environment (that is, greater numbers of stakeholders involved in decisions) in which to make local decisions about curriculum and instruction than their counterparts in other nations. This complexity leads to a decisionmaking process that is negatively related to achievement. Although these are just some of the possible causes of the poor international performance of American disadvantaged students in high school mathematics, they illustrate the massive challenge before the United States to improve education for a group of students and families that has continually been left behind.

Implications for American Educational Policy

Over the past century, the image of the American public high school has gone from an international model of effective secondary education to a symbol of the United States' educational crisis. How much these contrasting images match reality is not easy to determine. From an international perspective, the American high school clearly is not as negatively extreme in its practices and environment as some of its critics have suggested. However, American high schools fail to do some things high schools in other systems routinely do to advance mathematics achievement for students.

In considering U.S. schools internationally, one must keep in mind that significant global forces are at work to lessen cross-national differences in how schooling operates. The current global trend is toward convergence among structure and operation of the world's school systems. For example, in a 2001 paper, Gerald K. LeTendre, Motoko Akiba, Brian Goesling, Alexander W. Wiseman, and I use the TIMSS case-study transcripts and teachers survey and find substantial isomorphism among aspects of teaching in Germany, Japan, and the United States.²⁹ Another paper shows that a widespread single pattern of smaller school effects and larger family background effects on student mathematics achievement across all TIMSS nations is in part a function of the homogenizing of schooling through increasing state support of education.³⁰

The analyses reviewed here suggest several conclusions about educational reform and the state of the American high school. First, that no single silver bullet can solve U.S. education problems is certainly true, if somewhat incomplete. Analyses of international data over the past several years suggest that a whole case of silver bullets may be required to make real changes in the international competitiveness of American secondary education. If concern about international rankings is not just rhetorical and the United States is truly dedicated to reforming mathematics and science education so that American students can be more competitive internationally, then educators and policymakers must realize that the international findings point to a need for a robust set of reforms instead of more singular approaches.

An example of this is the need to look beyond the messages about reform initially originating from TIMSS. A dominating policy message from earlier press releases and the earliest analyses of TIMSS is that the American mathematics and science curricula are broken and need reforming. I would cer-

tainly agree that the United States should reform its curricula along more effective standards; there is nothing to disagree with about this. And certainly comparative analysis has helped reveal alternative models for curricula. But subsequent analyses of TIMSS show how complex the whole picture is. Reforming key operational pieces of the American school such as the intended curricula or the accountability of instruction without more effective distribution of basic instructional resources, particularly to the schools that need resources the most, will not lead to significant national improvement. The same can be said of the current major reform idea of accountability and local control of instruction and learning. These are clearly good ideas, but to launch them without coordination with other reforms, such as resource and teacher quality improvement, will most likely not achieve much national change. Certainly the desperate conditions of education for disadvantaged American students will necessitate simultaneous reform on a number of fronts.

This kind of reform is not easy to accomplish. A key difficulty of American education policy reform is the task of making broad change palatable to the multilayered structure of districts, states, and federal government, intermixed with a vast array of powerful interest groups. Consequently it is necessary to push hard for any particular reform, and often this leads to simpler, stripped-down messages. American educational policy recommendations are aptly named reform movements, akin to political and social movement campaigns. To change policy in such an environment, simplified singular reforms, instead of more complex multiple reforms, have the best chance of successfully running the political maze of the U.S. educational system. This presents a formidable challenge to getting American education to adopt a broad integrated approach to reform suggested by international findings.

Looking across these analyses of TIMSS and other similar international data gives one concern about finding clear ways to make national policy. Some of the surprises found in research studies illustrate this best. For example, finding that large amounts of shadow education are remedial in most nations and that a positive relationship does not exist between substantial outside education and national achievement came as a surprise given the volume of speculation about the positive effects of enhancement shadow education in high-performing nations such as Japan and South Korea. Or, finding that widespread daily homework in a system is perhaps an indicator of weak instruction and curricular implementation seems at first a paradox, but upon further reflection makes sense. National teaching corps that use

homework judiciously in the teaching process and less routinely (that is, nonritualistically) reflect a number of positive things about a national education system. And, finding that national rates of school disruption and everyday violence are more a function of the overall quality of schooling and not a function of adult violence in a nation suggests a host of policy implications that are not usually considered in terms of curbing school violence and student disruption.

Nations are nations, schools are schools, and students are students; causal factors at one level do not necessarily or automatically transfer to other levels. The full-scale implementation of secondary education in nations is a complex phenomenon, influenced by more than the simple linear aggregation of school-level or student-level factors. All levels of the system share processes, but the increasing organizational complexity as one moves organizationally further from the student adds entirely new dimensions that have an impact on the overall practice of education. The seemingly intractable problem of limited educational opportunities and poor academic outcomes of economically and racially disadvantaged students in the United States is a prime example of how elusive systemic solutions are. But at the same time, the United States has seen other nations do significantly better with their disadvantaged students. If nothing else, this provides some hope for more systemic national reforms.

Seen from an international perspective, the American high school and what goes on there suggest that education policy cannot be easily separated from other sectors of policy. For example, using the TIMSS data, Gillian Thompson shows that among European nations cross-national variation in family policy influences the detrimental effects of single-parent families on achievement.³¹ Nations with more progressive economic family assistance policies tend to have a weaker connection between family problems and mathematics achievement. This is preliminary work that needs more development, but along with the other analyses presented here it suggests that broader policy reforms, such as coordinating education and family policy, will be needed to make significant national progress in secondary education of all American students.

Comment by Jaekyung Lee

David P. Baker makes a significant contribution to the understanding of cross-national education differences by synthesizing many studies that he and his colleagues have done with the Third International Mathematics and Science Study (TIMSS) data and by drawing policy implications from that synthesis. This paper not only addresses the issue of poor performance of the average American student from an international perspective, but also pays attention to the poorer achievement of disadvantaged students in the United States. Baker's paper provides a comprehensive look into school effect by combining a complex array of the TIMSS data on school input, process, context, and outcomes. However, the paper also reveals the limitations of the current TIMSS data, as well as limitations in Baker's studies as they relate to critical policy questions.

Issues of Selection for Review and Comparison

In reviewing and synthesizing past TIMSS studies, Baker selects almost exclusively his own studies without referencing other studies on the same topic. Providing a comprehensive and balanced literature review, including earlier international studies using non-TIMSS data as well as the studies of TIMSS data by other researchers, would have created a scholarly framework within which his findings could be better evaluated. In the comparative research field, different studies often produce findings that are not always congruent with each other. Much depends on which countries and variables are selected and how they are analyzed and compared.

Who is being compared with whom is critical in this kind of comparative study. The validity of selection and comparability is not discussed in this paper. Baker is not consistent in his selection of countries for comparison. In the first part of his paper, when examining average achievement, Baker includes all TIMSS countries for comparison. But in the second part of his paper, where he examines the achievement of disadvantaged students, Baker selects only developed, wealthy countries. Moreover, he uses both eighth- and twelfth-grade data for the former comparison but only twelfth-grade data for the latter comparison. It is not clear why he chose to use different sets of countries or grades for comparisons and how such different selection criteria might have influenced his findings and conclusions.

Missing Variables and Measurement Issues

The studies that Baker reviews use several indicators to measure school effects, but some important variables are missing in those analyses. To begin with, information on the rigor of the curriculum is lacking. Based on the TIMSS teacher survey results, Baker points out that the average U.S. eighth-grade math class has more coverage and repetition of topics and that these variables are not related to achievement variations across the TIMSS countries. Without considering other key instructional variables, however, how the breadth of instruction alone can be related to achievement outcomes is not clear. The level of implemented curriculum should have been considered. TIMSS curriculum studies showed that the U.S. curriculum is not only less focused but also less advanced.³² Further, Baker's comparison of implemented curriculum is limited to the eighth-grade data, and inferring that the American high school is not much different from an international average based on the eighth-grade level data is misleading. Simply comparing the national averages also conceals substantial variations in the type of courses offered by American high schools and taken by their students.

Teacher quality and training are another important set of missing variables in Baker's review. He acknowledges that the TIMSS data provide little information on the training of teachers. In fact, teacher quality can be the most important factor that determines the quality of instruction. Even when teachers say that they cover the same thing to the same extent, their differences in content and pedagogical knowledge should result in significant variation in the quality of teaching and learning.

In addition to these kinds of key schooling variables, much broader cultural and institutional differences must be considered. East Asian countries such as Japan and South Korea could not have performed best without their traditional cultural influences: high expectations for academic achievement, high aspiration for college education, high level of parental engagement and support for learning, and highly homogeneous value system. These cultural forces interplay with or are reinforced by centralized institutional rules and arrangements that restrict student choices, impose fierce competition for a common goal (for example, entrance into top-tier universities), and reward academically successful students.

Private tutoring in Japan and South Korea reflects these forces and plays an important role in boosting academic achievement.³³ Although Baker argues that this out-of-school education variable (shadow education) gener-

ally does not explain cross-country variation in achievement, the variable measured by students' self-reported frequency and amount of after-school lessons may not capture the level of rigor and intensity that are strongly influenced by cultural and institutional forces. In a similar vein, his argument that high-stakes exams have no effect on national achievement needs further investigation. His analysis does not differentiate how high the stakes are and how competitive the exams are.³⁴

In his analysis of equity, Baker uses two indicators of socioeconomic status (SES): mother's education and single-parent household. He argues that these family characteristics often lead to a disadvantaged home life for many students. One test of this argument is whether these are good predictors of student achievement. Common sense says yes. However, an analysis of the TIMSS data by D. Koretz and his colleagues shows the insignificant influence of mother's and father's education on student achievement in the United States.³⁵ They also find that the effects of mother's education in TIMSS were less than half the size compared with the results from the National Education Longitudinal Study. The reason for this is unknown, but the finding raises doubt about the reliability of students' self-reported parental education variable in TIMSS. Checking other indicators of SES such as the number of books available at home would be helpful.

Issues of Data Analysis and Potential Biases

Ecological fallacy means that the between-country relationship of variables cannot be used to suggest that the within-country relationship would be the same. In addition to cross-national analysis, unique predictors of academic achievement need to be identified country by country. For example, T. N. Postlethwaite observes that the relationship between whole class teaching and achievement is positive at the between-country level, whereas most countries have no relationship and some countries even have a negative relationship at the within-country level.³⁶ Baker also points out that causal relationship at one level does not necessarily transfer to other levels. This raises a question as to what would be the most appropriate level or unit of analysis to guide policymaking. The answer depends on at which level of a school system a given schooling variable can be best manipulated. Can highly decentralized countries such as the United States manipulate daily assigned homework? If the classroom is regarded as the most appropriate locus of control over homework, then classroom-level analysis of the rela-

tionship between homework and achievement should provide the most meaningful and informative results. This issue also relates to the question of whether the United States can be fairly compared with other countries that have a centralized national school system with more control over key schooling variables.

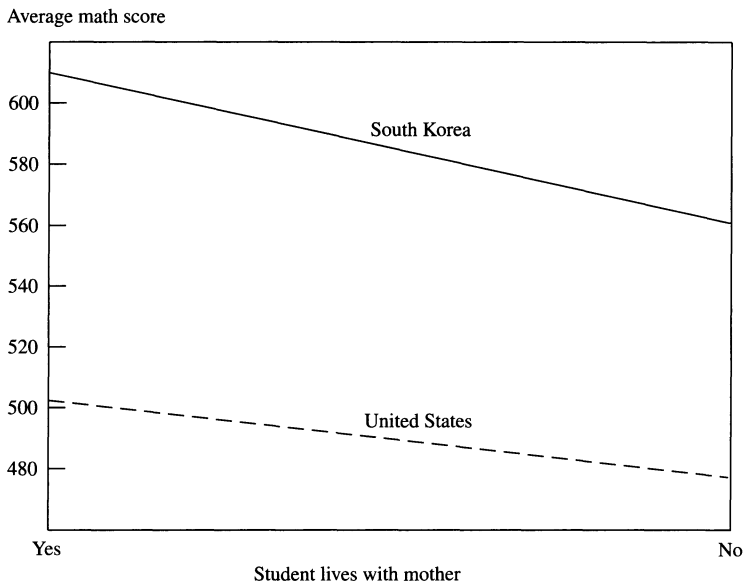
Furthermore, drawing causal inferences from cross-sectional data analysis is difficult. For example, Baker observes that the relationship between homework assignment and achievement is negative. Even though the analysis was done at the classroom level, the result may be interpreted in a different way: When students were not performing well, teachers were more likely to assign homework. The same can be said of the relationship between shadow education and achievement: Lower-performing students were more likely to seek private tutoring help. Because of these limitations, Baker's analysis of the TIMSS data does not provide clear answers about school effect. Despite the cross-sectional nature of TIMSS data, more analysis of achievement gains, instead of achievement status, is needed. Why does U.S. students' achievement deteriorate from the eighth grade to the twelfth grade relative to other countries?³⁷ Which countries improved student learning more than others at the eighth-grade level and why? By addressing these questions, observed achievement gains could be better related to education policies and policy effects could be better assessed.

Achievement Gap and Inequity

Baker's comparison of disadvantaged students' achievement deserves further analysis. The meaning of disadvantage may not be comparable from country to country. Do single-mother households in Europe have the same level of socioeconomic disadvantage as their counterparts in the United States? Even within the United States, disadvantage has several dimensions and the indicators may change as population changes over time. For example, students from a non-English language background have grown rapidly in the United States, and they face special disadvantages for schooling that other disadvantaged groups do not. More discussion about the nature and type of social or educational disadvantage in each country is needed.

Contrary to a popular belief, the U.S. distribution of academic achievement does not show any greater inequality than does an average country. In terms of variability within country as measured by standard deviation, U.S.

Figure 1. Average Eighth-Grade Math Scores by Living with Mother in South Korea and the United States

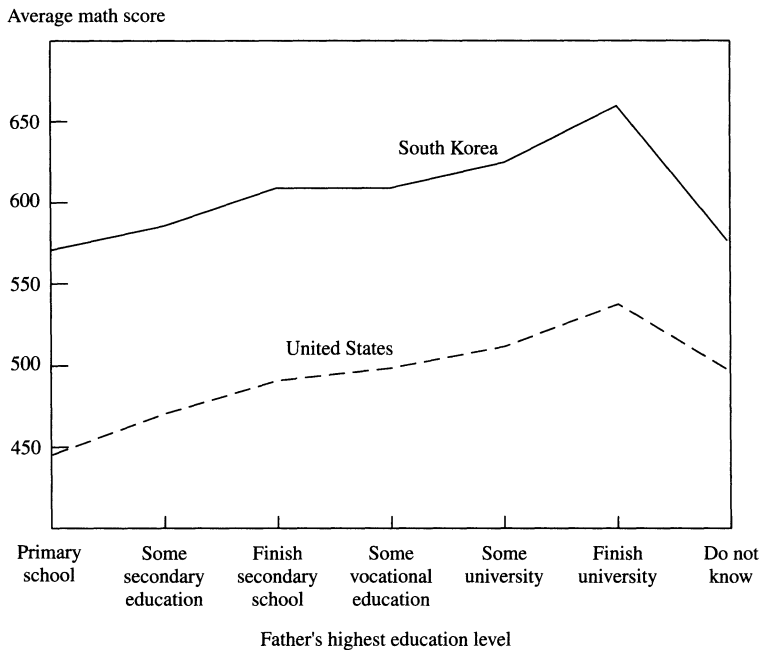


eighth graders are slightly above the international mean in mathematics and the United States does not have a large achievement gap.³⁸ It would be more meaningful to compare the achievement gap instead of the group average among the countries. How big is the gap between students from single-parent homes versus other students in the United States? How does the gap in the United States (not the average score of single-parent students) compare with the corresponding gap in other countries? For example, figure 1 and figure 2 show that the TIMSS eighth-grade math achievement gap between disadvantaged students and others in the United States, based on Baker's definition, is not different from South Korea. It would be more meaningful to relate the measure of the resource inequity to measures of achievement gap, instead of to the achievement of disadvantaged students.

Conclusion

The central question raised in Baker's paper is whether the United States can improve its performance by imitating what higher-performing countries do. Here performance concerns include both academic excellence and

Figure 2. Average Eighth-Grade Math Scores by Father's Highest Education Level in South Korea and the United States



equity. The former is usually measured by the level of average achievement; the latter, by the size of the achievement gap. Full agreement has not been reached on what factors account for the higher average achievement of countries such as Japan and South Korea. Moreover, consensus has not been found on whether the United States has more or less equitable distribution of achievement.

What seems obvious, though, is that the United States has moved in the direction of following those higher-performing countries' path during the last two decades. However, what many people may not realize as well is that those higher-performing East Asian countries followed the opposite path (for example, destandardization, diversification, decentralization) during the same period of time, and the Eastern and Western school systems may become more alike over time.³⁹ U.S. educational reform goals must be reconsidered. Educators and policymakers should be thinking about ways to maintain the strengths of the American school system as much as ways to fix its weaknesses. A balancing act is in order.

Comment by Stephen P. Heyneman

Cross-national education survey research was first just an experiment, born by a chance visit of Torsten Husen from the University of Stockholm to the Comparative Education Center at the University of Chicago in the mid-1950s. There, Husen met C. Arnold Anderson, Mary Jean Bowman, and Benjamin Bloom, whose view was that the whole world should be seen as a single educational laboratory. From this meeting emerged the International Association for the Evaluation of Educational Achievement (IEA), which, for diplomatic reasons, was managed from Sweden. First results appeared in 1964. Since then, thirty-three cross-national studies have been conducted, twenty-nine of which were associated with IEA.

From the beginning a myriad of problems emerged—the logistics of such a massive enterprise, the complexities of agreeing on common definitions, methodologies, sampling and data management. These problems have been the focus of a great deal of official and unofficial assessments of the state of the art, which helped to generate a number of significant improvements in the standards expected for cross-national surveys.⁴⁰ In spite of this progress, doubts and skepticism remain.⁴¹ These have generated carefully constructed replies.⁴²

Early reviews of cross-national policy implications focused attention on better understanding of the generalizations associated with the influences on academic achievement. One of the most persistent generalizations is that the influence of the home is greater than the influence of the school itself.⁴³ Stephen P. Heyneman and William Loxley, however, found that the degree of this influence varies across nations and that the lower a nation's gross domestic product (GDP), the more influence the school seems to have.⁴⁴ This finding has been a principal rationale for the investment in school quality by the World Bank, United States Agency for International Development (USAID), United Nations Educational, Scientific, and Cultural Organization (UNESCO), and many other development assistance agencies. Although recent reanalyses have challenged the strength of the earlier findings, the current conclusions are that the influence of socioeconomic status on achievement is by no means uniform across nations, age or grade levels, gender, and subject matter.⁴⁵

More recent policy reviews have tried to speculate on the meaning of these cross-national projects on local education policy within the United

States. William H. Schmidt and his colleagues, for instance, helped Americans focus on the weaknesses of having a splintered curriculum in which a scattering of topics is presented to students with insufficient attention to the progression, sequencing, and review found in more academically successful nations.⁴⁶

David P. Baker has used cross-national studies to generate novel hypotheses across a wide variety of education policy characteristics. He finds that American school systems distribute school resources less fairly and are less able to educate the most disadvantaged students by comparison to other countries. He also uses cross-national studies to challenge common assumptions—for instance, that American schools are more likely to experience violence and classroom disruption. He finds that more homework is often a proxy for poor educational quality and that nations with effective family assistance policies are better able to overcome the handicap of social poverty. And he uses these cross-national studies as the basis for speculating that the reason American school systems have such significant policy difficulties stems from their public governance in which elections for education posts determine policy direction. Because political campaigns have a single and simplistic focus (sex education, get God back in school, more time on task, class size, and so on) that attracts voters for school board elections, the nature of American education reform seems to be as fractured as the sequencing in its curriculum.

The problems in Baker's paper are few, but not without importance. Baker seems to assume that remedial work (the main focus of homework) is somehow less than useful. The term *remedial* seems to imply something pejorative. But in any survey the interpretation of what remedial means may vary. It may imply work to make up for previous gaps in understanding, or it may imply work on the essential skill mechanisms within a curriculum. For instance, it may include craftsmanship skills, something many schools in the United States lack.

For example, one school I visited recently in Russia displayed a large number of drawings in a hallway. All were considered excellent, but all were of one thing: a half-empty glass of water on a table. I asked why all the drawings were of the same thing. The answer was that they represented an essential exercise that all students (not just art students) must master before they can go on to more creative expression, and the display was supposed to highlight the differences in which creativity could be expressed within a restricted task.

I wondered how many schools in the United States would have taken such a mundane task to such an honorable level. I wondered whether the average level of drawing among students in the United States was adversely affected by the lack of attention to the mundane craftsmanship skills necessary to draw well. Could the tendency to use homework as remedial work be a sign that American teachers were attempting to lay the necessary groundwork for the craft of math and science?

In much of Baker's discussion of school governance, only two sets of actors are mentioned as having legitimate claims to schools. This is a serious mistake often made by American education policy analysts. The discussion of school choice is distorted by suggesting that it is dichotomous, that it is a tension between parent and state. Three different sources of legitimate claims normally are made on the governance of schools and school systems. They are the state, the parents, and the education profession.⁴⁷ The fact that the education profession is often left out of discussions about school choice in the United States may be a reflection of the weakness of the education profession in determining the future of American education. But just because it may be a weak source of governance within the United States, one should not assume that it is weak elsewhere.

Baker observes that the degree of state control differs from one country to another, and between local and national authorities, but he does not point out that the level of control by the education profession differs as well. Key to understanding education governance issues across countries is to understand better that legitimate control over schools and school systems is not binary. In many instances, understanding about what is possible in terms of governance, and characterization of countries as being more pro-parent choice as opposed to more controlled by state authorities, is flawed by such a simplistic binary model.

Torsten Husen, C. Arnold Anderson, Mary Jean Bowman, and Benjamin Bloom should be proud today. Almost fifty years after their meeting, their dream of using the world as a research laboratory to better understand the process of education has been realized. Cross-national projects, which attempt to calibrate the degree of curriculum coverage and the influences on achievement, are a normal, regular, and permanent part of the world of education.

Baker's paper represents the best of this genre of new and innovative interpretations of what Bloom had in mind when he thought of the world of education as a single laboratory. He did not think that it was uniform or that

identical principles or intervention strategies would emerge. Instead, he was energized by the possibility that all peoples could learn by looking at themselves through information systematically collected from around the world. This is what Baker's paper does so well.

Notes

1. David Tyack, *The One Best System: A History of American Urban Education* (Harvard University Press, 1974).
2. James Conant, *The American High School Today* (New York: Signet, 1959).
3. D. Baker and D. Holsinger, "Human Capital Formation and School Expansion in Asia: Does a Unique Regional Model Exist?" *International Journal of Comparative Sociology* (1996).
4. National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Education Reform* (Government Printing Office, 1983).
5. In the initial official reports on the Third International Mathematics and Science Study (TIMSS), the Department of Education included preliminary bivariate analyses. More complex analyses slowly emerged after these reports were issued.
6. For a description of the TIMSS study and data, see International Association for the Evaluation of Educational Achievement, *User Guide for the TIMSS International Database: Primary and Middle School Years*, edited by Eugenio J. Gonzalez and Teresa A. Smith (Chestnut Hill, Mass.: TIMSS International Study Center, 1997).
7. As described in the technical papers, all associations between reform factors and national achievement reported here have been verified using various appropriate multivariate modeling techniques including estimation by ordinary least squares regression, Hierarchical Linear Models, and structural equations.
8. K. Maguire and A. L. Pastore, eds., "Sourcebook of Criminal Justice Statistics, 1998" (www.albany.edu/sourcebook [1998]).
9. Department of Education and Department of Justice, *1999 Annual Report on School Safety* (Government Printing Office, 2000); and M. D. Hinds, ed., *Violent Kids: Can We Change the Trend?* (New York: National Issues Forums Institute and Public Agenda, 2000).
10. Motoko Akiba and others, "Student Victimization: National and School System Effects on School Violence in Thirty-seven Nations," *American Education Research Journal* (forthcoming).
11. M. Fernanda Astiz, Alexander Wiseman, and David P. Baker, "Slouching towards Decentralization: Consequences of Globalization for Curricular Control in National Education Systems," *Comparative Education Review*, vol. 46, no. 1 (2002), pp. 66–88.
12. Alexander W. Wiseman, "Principals' Instructional Management Activity: A Cross-National Analysis of How School Environments Influence What School Principals Do," Ph.D. dissertation, Pennsylvania State University, Educational Theory and Policy Program, 2001.
13. Gerald LeTendre and others, "Homework and the Quality of Education: A Cross-National Exploration," Working Paper (Pennsylvania State University, Education Policy Studies Department, 2002).
14. David P. Baker and others, "Worldwide Shadow Education: Outside-School Learning, Institutional Quality of Schooling, and Cross-National Mathematics Achievement," *Educational Evaluation and Policy Analysis*, vol. 23, no. 1 (2001), pp. 1–17.

15. National Research Council, *High Stakes: Testing for Tracking, Promotion, and Graduation* (Washington: National Academy Press, 1999).

16. For modest evidence of an effect, see J. Bishop, *Do Curriculum-Based External Exit Exam Systems Enhance Student Achievement?* CPRE Research Report Series RR-40 (University of Pennsylvania, Graduate School of Education, Consortium for Policy Research in Education, 1998).

17. Because chapter 6, "American High Schools in an International Context" by William H. Schmidt, in this volume compares in detail the American twelfth-grade curricula with other nations, I do not pursue all the intricacies of this topic here. W. Schmidt and others, *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics* (Netherlands: Kluwer Academic Publishers, 1997).

18. Gerald LeTendre and others, "Classroom Implementation of National Curricula and Cross-National Patterns of Achievement," Working Paper (Pennsylvania State University, Education Policy Studies, 2002).

19. Brian Goesling and others, "The Institutional Context of Educational Stratification: Comparing School Resource Inequalities within Thirty-five Countries," paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, April 2002.

20. My colleagues and I are currently undertaking this kind of analysis for the twelfth-grade sample. Preliminary results indicate a similar trend cross-nationally and for the United States.

21. W. Wilson, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy* (University of Chicago Press, 1987).

22. Glenn Firebaugh, *The New Geography of Global Income Inequality* (Harvard University Press, 2002).

23. J. Kozol, *Savage Inequalities* (Crown Publishers, 1991).

24. George Farkas, "Racial Disparities and Discrimination in Education: What Do We Know, How Do We Know It, and What Do We Need to Know?" workshop paper on Measuring Disparities in Education, National Research Council, Division of Behavioral and Social Sciences and Education, Committee on National Statistics, Washington, D.C., July 1, 2002.

25. While the United States has a high stop-out rate (around 20 percent) by age eighteen, it offers a number of ways to obtain a high school diploma. As a result, the United States has one of the highest high school completion rates in the world in cohorts aged twenty-four to twenty-eight.

26. David P. Baker, Gerald LeTendre, and Brian Goesling, "Educational Resource Inequalities among American Schools and Cross-National Differences," Working Paper (Pennsylvania State University, Education Policy Studies Department, 2002).

27. Similar results are found for other indicators of student disadvantage.

28. I took the quartile of schools that had the largest proportion of students with mothers with less than high school education.

29. Gerald K. LeTendre and others, "Teachers' Work: Institutional Isomorphism and Cultural Variation in the U.S., Germany, and Japan," *Educational Researcher*, vol. 30 (2001), pp. 3-15.

30. David P. Baker, Brian Goesling, and Gerald K. LeTendre, "Socio-Economic Status, School Quality, and National Economic Development: A Cross-National Analysis of the 'Heyneman-Loxley Effect' on Mathematics and Science Achievement," *Comparative Education Review*, vol. 46, no. 3 (2002), pp. 291-313.

31. Gillian Thompson, "Does Family Policy Environment Mediate the Effect of Single-Parenthood on Children's Academic Achievement? A Study of Fourteen European Coun-

tries,” paper presented at the annual meeting of the Comparative and International Society, Orlando, Florida, March 2002.

32. What makes such diffuse curriculum worse is that it is less challenging. The topics being taught in U.S. eighth-grade math classrooms were at a seventh-grade level compared with other countries. See National Center for Education Statistics, *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context* (Department of Education, 1996); and W. H. Schmidt, C. C. McKnight, and S. T. Raizen, *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education* (Boston: Kluwer Academic Publishers, 1997).

33. See J. Lee, “Missing Links in International Education Studies: Can We Compare the U.S. with East Asian Countries?” *International Electronic Journal for Leadership in Learning*, vol. 3, no. 18 (1999), available at www.acs.ucalgary.ca/~iejll/volume3/lee.html [April 2002].

34. See M. A. Eckstein and H. J. Noah, *Secondary School Examinations: International Perspectives on Policies and Practice* (Yale University Press, 1993).

35. See D. Koretz, D. McCaffrey, and T. Sullivan, “Predicting Variations in Mathematics Performance in Four Countries Using TIMSS,” *Educational Policy Analysis Archives*, vol. 9, no. 34 (2001), available at epaa.asu.edu/epaa/v9n34 [April 2002].

36. T. N. Postlethwaite, “Overview of Issues in International Achievement Studies,” paper prepared for a lecture at Oxford University, Department of Education, 1997.

37. Another limitation of the Third International Mathematics and Science Study data is that the achievement test score is not on a developmental scale so an absolute amount of learning gain from the eighth grade to the twelfth grade cannot be examined. Moreover, differences in twelfth-grade students’ age level among countries make the comparison difficult. But comparison can be limited to countries that participated in both grades and have the same age level of students as the United States.

38. See also Koretz, McCaffrey, and Sullivan, “Predicting Variations in Mathematics Performance in Four Countries Using TIMSS.”

39. See J. Lee, “School Reform Initiatives as Balancing Acts: Policy Variation and Educational Convergence among Japan, Korea, England, and the United States,” *Educational Policy Analysis Archives*, vol. 9, no. 13 (2001), available at epaa.asu.edu/epaa/v9n13.html [April 2002]; and R. Rothstein, “Weighing Students’ Skills and Attitudes,” *New York Times*, May 16, 2001, p. A25.

40. I. Olkin and D. T. Searls, “Statistical Aspects of International Assessments of Science Education,” paper presented at the conference on Statistical Standards for International Assessments in Pre-College Science and Mathematics, Washington, D.C., 1985; D. Horvitz, “Improving the Quality of International Education Surveys,” draft prepared for the Board on International Comparative Studies in Education, 1992; E. A. Medrich and J. E. Griffith, *International Mathematics and Science Assessments: What Have We Learned?* (Department of Education, Office of Educational Research and Improvement, 1992); H. Goldstein, *Interpreting International Comparisons of Student Achievement* (Paris: United Nations Educational, Scientific, and Cultural Organization, 1995); T. N. Postlethwaite, *International Studies of Academic Achievement: Methodological Issues* (University of Hong Kong, 1999); Robert L. Linn, “The Measurement of Student Achievement in International Studies,” in National Research Council, *Methodological Advances in Cross-National Surveys of Educational Achievement* (Washington: National Academy Press, 2002), pp. 27–58; Ronald R. Chromy, “Sampling Issues in Design, Conduct, and Interpretation of International Comparative Studies of School Achievement,” in National Research Council, *Methodological Advances in Cross-National Surveys of Educational Achievement* (Washington: National

Academy Press, 2002), pp. 80–117; Claudia Buchmann, “Measuring Family Background in International Studies of Education: Conceptual Issues and Methodological Challenges,” in National Research Council, *Methodological Advances in Cross-National Surveys of Educational Achievement* (Washington: National Academy Press, 2002), pp. 150–98; Robert E. Floden, “The Measurement of the Opportunity to Learn,” in National Research Council, *Methodological Advances in Cross-National Surveys of Educational Achievement* (Washington: National Academy Press, 2002), pp. 231–67; Marshall S. Smith, “Drawing Inferences for National Policy from Large-Scale Cross-National Education Surveys,” in National Research Council, *Methodological Advances in Cross-National Surveys of Educational Achievement* (Washington: National Academy Press, 2002), pp. 295–321; Brian Rowen, “Large-Scale, Cross-National Surveys of Educational Achievement: Promises, Pitfalls, and Possibilities,” in National Research Council, *Methodological Advances in Cross-National Surveys of Educational Achievement* (Washington: National Academy Press, 2002), pp. 321–53; National Research Council, *Summary Report of the Conference on October 16-17, 1985 (Draft)*, Committee on National Statistics, Commission on Behavioral and Social Sciences and Education (Washington: National Academy Press, 1985); National Research Council, *A Framework and Principles for International Comparative Studies in Education*, Board on International Comparative Studies in Education, edited by Norman M. Bradburn and Dorothy M. Gilford, Commission on Behavioral and Social Sciences and Education (Washington: National Academy Press, 1990); National Research Council, *International Comparative Studies in Education: Descriptions of Selected Large-Scale Assessments and Case Studies*, Board on International Comparative Studies in Education, Commission on Behavioral and Social Sciences and Education (Washington: National Academy Press, 1995); and M. O. Martin, K. Rust, and R. J. Adams, *Technical Standards for IEA Studies* (Amsterdam, Netherlands: International Association for the Evaluation of Educational Achievement, 1999).

41. I. Rotberg, “I Never Promised You First Place,” *Phi Delta Kappan* (December 1990), pp. 296–303.

42. N. Bradburn and others, “A Rejoinder to ‘I Never Promised You First Place,’” *Phi Delta Kappan* (June 1991), pp. 774–77.

43. James S. Coleman and others, *Equality of Educational Opportunity* (Department of Health, Education, and Welfare, Office of Education, 1966).

44. Stephen P. Heyneman and William Loxley, “The Effect of School Quality on Academic Achievement across Twenty-nine High- and Low-Income Countries,” *American Journal of Sociology*, vol. 88, no. 6 (May 1983), pp. 1162–94.

45. David P. Baker, Brian Goesling, and Gerald K. LeTendre, “Socioeconomic Status, School Quality, and National Economic Development: A Cross-National Analysis of the ‘Heyneman-Loxley’ Effect,” *Comparative Education Review* (forthcoming).

46. William H. Schmidt and others, *Why Schools Matter: A Cross-National Comparison of Curriculum and Learning* (New York: Jossey-Bass, 2001).

47. Amy Gutmann, *A Democratic Education* (Princeton University Press, 1999).