Measuring Up to the Competition:

A Few Hard Questions

By Iris C. Rotberg

WHAT'S WRONG WITH COMPARING AMERICAN STUDENTS' TEST SCORES TO THOSE OF STUDENTS IN OTHER COUNTRIES? PLENTY, ACCORDING TO IRIS ROTBERG. SHE SAYS, "THE DIFFICULTY IS NOT IN DEVISING ELEGANT STATISTICAL DESIGNS BUT IN CARRYING THEM OUT IN THE REAL WORLD." MOREOVER, A PREOCCUPATION WITH TEST COMPARISONS LEADS US TO "RECOMMEND SOLUTIONS THAT ARE IRRELEVANT AT BEST AND OFTEN ARE COUNTERPRODUCTIVE." HERE, SHE ANSWERS QUESTIONS ABOUT INTERNATIONAL TEST-SCORE COMPARISONS AND AMERICAN COMPETITIVENESS.

Iris C. Rotberg, a senior social scientist with RAND's Institute on Education and Training in Washington, D.C., directs research on federal policy options for improving the education of low-income students. Her article, "I Never Promised You First Place," was published in the December 1990 issue of *Phi Delta Kappan*. These comments are adapted from informal remarks she delivered at a forum on the value of international assessments presented at the Office of Educational Research and Improvement, U.S. Department of Education, on March 23, 1992.

International comparisons of science and mathematics achievement reportedly show that the U.S. education system is failing. That conclusion, which I believe is based on studies that are seriously flawed, has deflected our attention from real and difficult problems. So, too, has the rhetoric about U.S. competitiveness, which is assumed to be a deficiency attributed primarily to the quality of science, engineering, and technology education. I also question the assumption that our education problems can be resolved without attention to the underlying conditions of poverty. These myths, while rhetorically satisfying, inevitably lead us to recommend solutions that are irrelevant at best and often are counterproductive to resolving or even addressing our most important problems. My conclusions are presented best, perhaps, by posing a set of questions that I am frequently asked about these issues.

Since the international test-score comparisons all seem to produce similar negative findings, doesn't that mean that there must be some underlying validity to them?

No, it means that they all have the same shortcomings. It is simply not feasible to control for the major societal differences among nations. First, the students represented in the test comparisons are much more highly selected in some countries than in others. Second, some countries, like the United States, have a relatively high proportion of low-income students who are in school and taking the tests. Countries with substantial proportions of low-income students taking the tests tend to score lower than countries with less poverty or than those whose low-income students are not tested simply because they are not in school. Third, there are differences in curriculum emphases among nations that contribute to the relative rankings. If students have never studied calculus, we can be sure they will not do well on a calculus test. We don't need to administer an international test to tell us that. While there is room for debate about whether a higher proportion of U.S. high school students should take calculus, this issue cannot be resolved on the basis of test scores of students who have never taken the subject.

Why did you question the findings of the international comparisons in the first place?

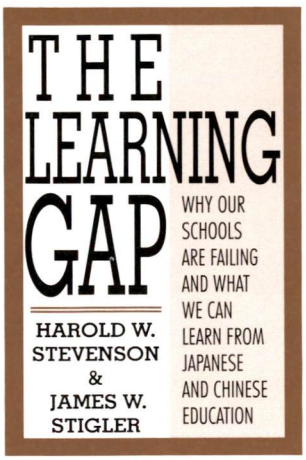
The test-score results could not be explained by differences in the quality of education. For example, there were major reversals of rankings between higher and lower grades in Hungary, England and Wales, British Columbia, Japan, Hong Kong, the former Soviet Union, Slovenia, and the United States. Consider the results of a recent assessment of mathematics students in Hungary and England. Hungary ranks near the top in the eighth-grade comparisons. Not surprisingly, by the twelfth grade, when Hungary retains more students in mathematics than any other country, Hungary ranks among the bottom countries. Have Hungary's schools gone downhill between the eighth and the twelfth grades? Or is it simply a matter of more stu-

students, and the gap in their performance widened the longer they were in school.

The decade of the 1980s was a time of heightened interest in American education, and everyone from the President and the governors to local school boards asked how improvements could be made. To evaluate whether the scores in reading improved over this period of concern and innovation, we went back in 1990 to the same schools we had visited a decade earlier. We tested several hundred fifth graders in each city with the same test that had been used in 1980. The mean score had increased by one point—hardly a notable improvement. The fifth graders also were given the reading test their counterparts had been given in 1980. Rather than showing improvement, their status relative to their Chinese and Japanese peers had deteriorated, as indicated in Figure 2.

The Asian Way

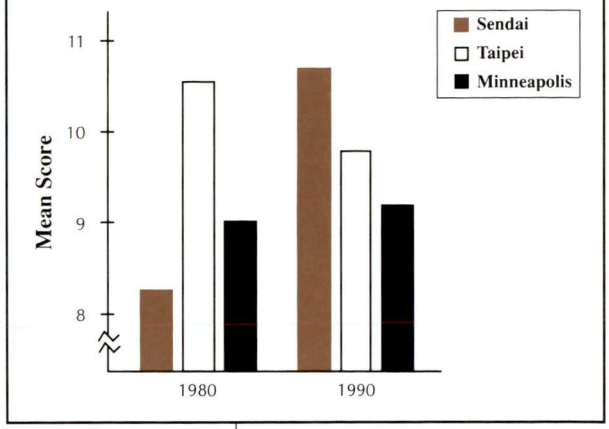
Why do Asian students do so well? There are many reasons for this success, but underlying their achievements is the firm belief that the avenue to success lies, as it has for hundreds of years in Asian cultures, in becoming a learned person. This belief in the vital importance of education, along with the belief that all students are capable of learning if they are taught well and are diligent, helps us understand many Asian practices. Influenced by ancient Confucian teachings about human malleability, Asian teachers work long hours trying to help their students. Asian parents make sacrifices to provide the types of home environment and personal support they believe to be necessary for academic success. And Asian students accept the need to study hard and pay close attention to their school work.



This optimistic view of how dedication and effort may lead to success was revealed in a conversation one of my colleagues recently had in Taiwan. She asked a group of teachers what subject they most liked to teach; she knew from previous experience that they would say mathematics. Why did they like to teach mathematics? "Because there are such individual differences among the students," they answered. Puzzled, she persisted in her questioning. Why should this make them favor mathematics?

"Because," they answered, "it is so challenging and so rewarding to bring all children up. Most students eventually learn to read, but only through a good teacher can some children learn mathematics."

Figure 2



The Role of Technology

Visitors to Asian classrooms are surprised to find no computers or even hand calculators. The most advanced piece of equipment found in the typical Taiwanese elementary school classroom is an overhead projector; in Japan it is a television set. Chinese classrooms are lucky if they have either.

We recently asked nearly 4,000 eleventh graders in Taipei, Sendai, and Minneapolis about their use of hand calculators in mathematics classes. Only 10 percent in Sendai, 42 percent in Taipei—but 88 percent of the students in Minneapolis—said they used hand calculators. Less than 1 percent of the Chinese and Japanese students, compared to 58 percent of the American students, said they always used them in their mathematics exams. The use of computers in learning school subjects paralleled that for hand calculators, except that only 15 percent of the American students used computers for this purpose.

The low reliance on technology in Asian classrooms is determined partly by economics. With the continued growth of Asian economies and greater investment of funds in education, it is likely that the use of technology will increase. One of the fascinating studies in the coming years will be to see what happens to the already outstanding performance of Asian students when advanced technology is used more frequently in their schools.

Then how will American students measure up? ◀

The most complete summary of the results of all the large international studies is found in "International Mathematics and Science Assessments: What Have We Learned?," published by the U.S. Department of Education (1992).

See "Contexts of Achievement: A Study of American, Chinese, and Japanese Children," in *Monographs of the Society for Research in Child Development* (1990), for the most succinct presentation of findings by H.W. Stevenson, S.Y. Lee, C. Chen, J.W. Stigler, C.C. Hsu, and S. Kitamura.

AIT photo

dents, and therefore lower scores? England, by contrast, scores in the bottom half in most of the eighth-grade comparisons, but ranks among the top countries by the twelfth grade, when only a highly selected group of students there is tested. Of course, this type of sampling problem is not limited to international comparisons. Within the United States, the relative rankings of states on average SAT scores are also a reflection of the proportion of students who take the test. The states with the highest proportions of students taking the SAT tend to have the lowest average SAT scores.

With all our expertise in statistics and sampling design, can't we simply improve the validity of the international comparisons?

No, the fact is that we have had expert statisticians working on the problems for the past 30 years. The difficulty is not in devising elegant statistical designs but in carrying them out in the real world. The problems in making these comparisons are endemic to all of the studies, including the most recent Educational Testing Service (ETS) study.* ETS went out of its way to point out these problems and advised in its press release against ranking the countries. But more important, would our children's education improve if we established rigid international controls on each nation's sampling design, located out-of-school (or homeless) children and tested them on science and mathematics, or controlled for tracking or relative socioeconomic status? And even if we did so, what is the chance that the test-score differences could be attributed to the quality of each nation's education system?

Does it matter if we exaggerate the problem in the United States, when we all agree that science education can be better than it is?

Yes, it does matter. The rhetoric is not supported by the facts. We incorrectly assume that adverse test-score differences mean that our schools, or our parents, or our students, or our scientists, or our research institutions have failed. I am particularly concerned about proposed remedies based on misleading test-score differences—for example, raising course and graduation requirements—without doing anything about the vast financial differences between rich and poor school districts. Those requirements will do more harm than good. We are likely to screen out of the education system precisely those students who already receive the lowest-quality education. Certainly, we will reduce their graduation rates and subsequent employability and earnings. We will end up with a so-called meaningful high school certificate, but fewer students will receive it. What will the others do for a living?

Moreover, for all students, we are in danger of placing increased emphasis on rote learning measured by multiple-choice tests, and less emphasis on the type of curriculum changes and teaching practices that would focus on an understanding of basic scientific and engineering concepts and research methods. It is unlikely that memorizing facts that can be readily assessed on standardized tests will encourage greater numbers of high-achieving students to become scientists and mathematicians. Nor is it likely to give young people who do not attend college the skills

***Two reports of the International Assessment of Educational Progress: "Learning Science" and "Learning Mathematics," Educational Testing Service, Princeton, New Jersey, February 1992.**

they will need to compete in a world requiring ever greater technological skills.

Do you believe that we can learn something from other nations' education systems or teaching practices?

Of course. However, the challenge is to identify those practices that realistically can be replicated in the United States. While comparative international studies may provide some insights, relatively few produce findings that can be readily transferred from one nation to another. In most cases, it would involve a basic restructuring of a nation's social, cultural, and political institutions. It would involve

We cannot—nor would it be wise to—superimpose changes in education outside the context of a country's cultural and political environment.

changes in some rather fundamental aspects of our society: the respective roles of national and local governments in education, the role of the teacher in society, teachers' salaries, competitive sports in schools, summer vacations, our value system with respect to pluralism, open access to higher education across socioeconomic groups, the role of industry in vocational education and apprenticeship programs, and similar issues that each country looks at differently. The fact is we cannot—nor would it be wise to—superimpose changes in education outside the context of a country's cultural and political environment. But that is not a problem limited to education. For example, comparisons of industrial policy in the United States and Japan are just as complex. Government and industrial links differ so fundamentally between the two countries that, as a practical matter, it is not likely that the "lessons" learned in Japan can or should be transferred to the United States.

Aren't you too complacent about the problems in American science and mathematics education?

No. I just don't believe that the data support a conclusion that our schools have failed. However, I do think that we are far too complacent about the large proportion of our children who live in poverty, about the vast differences in educational resources between rich and poor school districts, about the rising costs of higher education, about reductions in the real value of student financial aid for low-income students, and about decreasing state expenditures for higher education—and what that does to student motivation. Unfortunately, we assume that schools can be improved with little attention to the underlying conditions of poverty and often hold schools accountable for fixing the fundamental problems in our society. My concern is that a focus on test scores deflects attention from what we can do to solve our real problems.

You talk about financial resources—isn't that just throwing money at the problem?

Not at all. We know that low-income and minority students, on average, have less opportunity to study science and mathematics than do other students. They also have less access to the most qualified teachers, to science laboratories, and even to up-to-date textbooks. We also know that there are large disparities in education spending between rich and poor school districts. In some states, per-pupil expenditures in affluent districts can be as much as two and one-half to three times as high as expenditures in the lowest-income districts.[†] If the amount of money spent on schools really doesn't make a difference, affluent parents haven't yet heard the message. If poor districts had as much funding as rich districts, they could reduce class size substantially, provide greater opportunities for individualized instruction, train teachers in new educational practices, incorporate technology into the instructional program, and still have enough left over to finance decent science laboratories.

Don't you believe that national testing would improve education for the students you are most concerned about?

On the contrary, I am concerned that there would be serious, negative consequences for them. Harold Howe II, senior lecturer in the Harvard Graduate School of Education and a former U.S. Commissioner of Education, described the potential impact of testing on students from low-income backgrounds who have major problems to overcome both outside and inside their schools. In a 1991 letter to John F. Jennings, general counsel for education to the House of Representatives Committee on Education and Labor, Howe writes: "Inside their schools, they are subjected to the effects of lower educational expenditures per student—larger classes, limited special services, decaying and inadequate facilities,

According to a 1992 report issued by the Organization for Economic Cooperation and Development, the United States is the wealthiest industrialized nation in the world. But in 1988, the United States spent only an average percentage of its gross domestic product on education, compared to 24 other countries studied. Denmark spent the highest percentage; Japan spent the lowest. Denmark reports a high school graduation rate close to 100 percent; Japan, nearly 90 percent. About 74 percent of U.S. 17-year-olds graduate.

—Editor's note

higher levels of teacher turnover and teacher absence, and numerous other signals that they are second-class citizens of the education system. To remind them with a new national test of these discouraging facts is not the best route to building their morale or their performance." And, of course, for all students there is the problem of increasing the emphasis on multiple-choice tests of isolated pieces of information.

Nor is the answer to national testing to develop innovative new examinations, such as performance assessments, essay exams, portfolio assessments. In how many years? At what cost? In recent testimony before the Committee on Education and Labor, several researchers estimated the cost of administering tests nationally in five subject matters in only three grades at more than $3 billion per year.[‡] By comparison, the entire Chapter 1 program, the largest federal program for elementary and secondary education, will spend $6.1 billion in the 1992–93 school year. I question whether test administration is the best use of scarce education resources.

What about American competitiveness? How can we compete in the global marketplace with the Japanese (the Koreans...the Germans...), if our students don't do better in these test comparisons?

Our problems in international competitiveness have little to do with the quality of science and engineering education. Rather, they are related to business practices, government policies, and the realities of a global economy—for example, exchange rates, the lack of incentives for industry to invest in long-term product development, the financial incentives that lead to off-shore manufacturing, differential wage rates among countries, differential government subsidies among countries, licensing practices, antitrust concerns, and the emphasis placed on military at the expense of civilian research. In the United States, approximately one-third of total expenditures (and two-thirds of federal expenditures) for research and development goes to defense. These are far more important explanations of the status of U.S. competitiveness than are rankings on international test comparisons.

A focus on test scores deflects attention from what we can do to solve our real problems.

Will the United States be first in the world in science and mathematics by the year 2000?

It all depends on the measure. If we choose our sample carefully, for example, only from the students attending the Bronx High School of Science, we will be first! More seriously, if we somehow measured our students' expertise in designing independent research projects, as demonstrated by the Westinghouse Science Talent Search, we would do quite well. The fact is our basic scientific research output is highly competitive right now—Nobel prizes, scientific publications, high-quality scientists and engineers. However, if we measure ourselves by international test scores, we will be far from first place. Far more important are other measures, such as those I just mentioned, and others: the vitality of our labor force, the employability and wages of those who do not attend college, and the quality of our workplace training in high-technology and information industries. Yes, our schools can be strengthened, but our success in doing so will require us to focus on a number of difficult public-policy issues rather than on test scores and rankings that tell us little about how to resolve or even identify the most serious problems. ◀

[†]William L. Taylor and Dianne M. Piche, "Shortchanging Children: The Impact of Fiscal Inequity on the Education of Students at Risk," a report prepared for the Committee on Education and Labor, U.S. House of Representatives, December 1991.

[‡]Daniel M. Koretz, George F. Madaus, Edward Haertel, and Albert E. Beaton, testimony before the Subcommittee on Elementary, Secondary, and Vocational Education, Committee on Education and Labor, U.S. House of Representatives, February 19, 1992.

Deus Machina

By Neil Postman

I

Once upon a time, in a land far away, disorder and fear plagued the people. Guns and cannons were everywhere, warring parties slaughtered each other by the thousands, and no soldier would venture into battle unless equipped with the most modern firearms. The gun makers of the land were powerful, skillful, and prosperous, for they not only made guns for their own people but sold them to foreigners as well. You could hardly travel anywhere in the cities or country without seeing a gun or hearing one, which is why the children slept fitfully, with fear in their hearts.

For almost one hundred years, this was the situation in that forlorn land. Then, gradually, the people began to wonder if they would not be better off without their guns. It is hard to know why this thought arose. But they were an intelligent people with strong and ancient traditions and a well-developed sense of civilized behavior. Perhaps that is why the soldiers announced that they did not really like guns, for there was little skill and no honor in killing a man with a gun. The politicians were forced to admit that guns were not necessary to protect the land from foreign invasion since their armies were large and loyal and had never forgotten how to use swords. Besides, no one had seriously tried to invade their land for as far back as anyone could remember. Then, too, everyone agreed that guns were ugly, hardly comparable to the elegant beauty of a well-made sword. And because the sword was so beautiful, it had a value far beyond its use as a weapon. It was a symbol of honor, piety, and courage. And everyone knew that there once was a time when swords were given as gifts to men of great character.

And so the politicians, the soldiers, the businessmen, and the plain folk decided it was best to give up their guns. This did not happen all at once, for people never agree to a thing one hundred percent. Some gun makers, for example, were not pleased until they realized that it was more fun and almost as profitable to return to making swords. And, of course, there were some soldiers who had never learned the art of swordsmanship and who worried about their future. But, eventually, people began to throw away their guns or sell them to the government, which was happy to destroy them. The government even paid the gun makers not to make guns, the way Americans pay their farmers not to grow food. In a short time, all the guns were gone. There were still wars, of course, for even in a fable the demons that make men war on each other cannot be wished away. But for two hundred years, the sweet song of the nightingale was never drowned by the retort of the rifle or the roar of the cannon. And the children slept peacefully, as they had done many years before.

II

I begin with a fable because it is the language traditionally used to imagine something that everyone knows cannot happen. But this fable has an ironic twist to it, since what it describes actually did happen. The faraway land is Japan, which in the sixteenth century was a world leader in the manufacture of matchlock guns and cannon, having been introduced to these technologies by European traders. Toward the end of that century, for reasons mentioned in the fable, the Japanese gave up their firearms and reverted to their traditional weapons. There were no guns in Japan until the mid-nineteenth century. All of this is meticulously documented in *Giving Up the Gun: Japan's Reversion to the Sword, 1543–1879,* by Noel Perrin (Boston: David R. Godine, 1979), who wrote the book for the best of reasons. He wanted to show that two of the fundamental beliefs of those who live in advanced technological cultures are at least slightly questionable. The first belief is that there are no circumstances in which the technological clock can be turned back. The second is that technology is autonomous and is therefore beyond the control of those who make machines or use them.

One must admit, of course, that there are not too many examples that refute either of these propositions. The case of the Japanese and their guns is one. The most well-known instance in the Western world occurred in England between 1811 and 1816, curiously, not far from the time when the Japanese resumed their use of guns. I refer to the much-maligned Luddite movement, the revolt of workers against the intrusion of machinery in the garment and fabric industry. The origin of the term "Luddite" is obscure, some believing that it refers to the actions of a youth named Ludlum who, being told by his father to fix a weaving machine, proceeded instead to destroy it. Perhaps. But it is certain that workers bitterly resented the fact that machinery had led to wage cuts, child labor, and the elimination of laws and customs that had protected skilled workers. Their discontent was expressed through the destruction of machines, and since then the term "Luddite" has come to mean an almost childish and naive opposition to technology. Of course, the historical Luddites were neither childish nor naive. Like the sixteenth-century Japanese, they were people trying desperately to preserve a world view that had given them a sense of worth and justice in an earlier time.

I bring all of this up because I believe there is something to be learned from these examples and these people. No, I do not expect all nuclear weapons to be dismantled or television to be blacked out or computers to be unplugged. There is a more realistic point to be made—an idea that was on the minds of the sixteenth-century Japanese and the nineteenth-century English