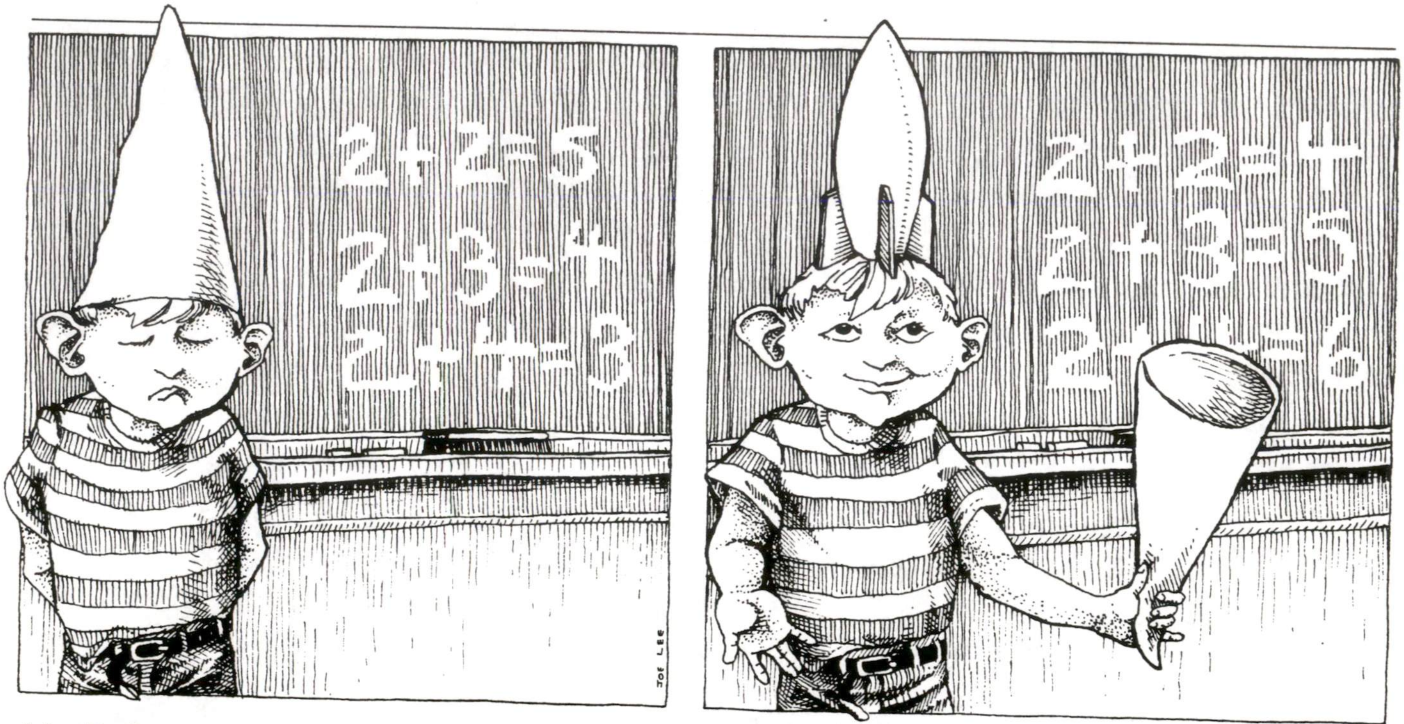
How Did All Those Dumb Kids Make All Those Smart Bombs?



Ms. Rotberg raises the question, Even if we could solve the methodological problems of international assessment, how would all these tests improve education for the children in any of the participating countries?

BY IRIS C. ROTBERG

EVER SINCE international comparisons of science and mathematics test scores began in the 1960s, Americans have believed the myth that U.S. students are outclassed by those in other nations. Yet, after almost three decades of apparent failures on international tests, we have somehow managed to maintain a level of research productivity that, by many measures of performance, is overwhelming. Some of us might wish that a greater proportion of this research were in the civilian sector, but that choice has nothing to do with the quality of our education system or with the expertise of our students.

IRIS C. ROTBERG is a program director with the Directorate for Education and Human Resources, National Science Foundation, Washington, D.C. The views expressed are the author's own and do not necessarily reflect the positions and policies of the National Science Foundation.

The apparent contradiction between low test scores and high research output would merely make interesting dinner conversation if U.S. officials were not taking the test score comparisons so seriously. As I noted in my December article, I am concerned that the public policy dialogue will continue to focus on test scores rather than on the far more important questions about our accomplishments – and our problems – in science and engineering education.[1] I think it would be useful to summarize that article's main points before commenting on the rejoinder by Norman Bradburn and his colleagues.

In my article I presented evidence that international comparisons of test scores

Illustration by Joe Lee

are highly misleading indicators of the effectiveness of a nation's education system or the quality of its scientific research.

First, the rankings of nations are biased because it is virtually impossible to control for the major societal differences among nations. Most nations retain a much lower proportion of students in the final years of high school than does the U.S. Therefore, a smaller proportion of their students – the highest-achieving students – are represented in the studies, compared to a much broader cross section of U.S. students.

I also noted that international test comparisons reflect the fact that the U.S. has a higher proportion of students living in poverty than many other industrialized countries – an unfortunate fact of U.S. society that will not be addressed by yet another round of international testing. Countries with substantial proportions of low-income students taking the test tend to score lower than countries with less poverty or countries whose low-income students are not tested simply because they are not in school.

In addition, I discussed curriculum differences from nation to nation that affect the test results. For example, advanced mathematics students in the U.S. are more likely to defer calculus until college than are their counterparts in many other countries.

The fact is that the quality of our scientific output and the skills of our science and engineering majors are extremely high. An analysis of mathematics scores on the Scholastic Aptitude Test shows that there is no shortage of highly qualified students. Indeed, the scores of top students have actually risen in recent years. In 1977 the 90th percentile score was 628; in 1986 it had risen to 642. And U.S. students continue to excel in competitions that reward excellence in independent research, such as the Westinghouse Science Talent Search.

The public perception that we are falling behind in science and mathematics is based on a narrow and highly questionable criterion. Moreover, our preoccupation with test comparisons may lead us to implement "solutions" that are at best trivial and may be counterproductive to addressing far more important problems.

For example, I pointed out in the December article that the most difficult challenge is not improving the quality of education for science and engineering majors, but providing a better education for the general student who will live in a world that requires ever-greater technological skills. It is unlikely that memorizing facts that can be readily assessed on international standardized tests will provide young people who do not attend college with the skills needed to compete in the workplace. Nor will test comparisons identify the reasons why U.S. scientific and technological advances are often not turned into products that international consumers want to buy. And international comparisons certainly will not focus attention on such important issues as the links between academic scientific research and industrial competitiveness, the lack of incentives for industry to invest in long-term product development, business practices that lead to offshore manufacturing, or the emphasis placed on military research at the expense of civilian research.

MY ORIGINAL article and the response from Bradburn and his colleagues agree on one point: international test comparisons still contain a number of methodological problems, despite researchers' best attempts to address them over a period of nearly 30 years. The main difference between the article and the rejoinder is in the conclusions drawn – especially with respect to the usefulness of international comparisons and the likelihood that the flaws in previous comparisons can be significantly reduced.

The rejoinder from Bradburn and his colleagues argues that the researchers are aware of the problems with the international comparisons. I agree. However, I was writng not about "awareness" of problems but about the likelihood of solving them. And I was writing not about expertise in devising elegant statistical designs but about the difficulty of carrying them out in the real world.

I would also note that Torsten Husén, one of the researchers who was cited in the rejoinder to support the theme of "researcher awareness," wrote an article in the March 1983 issue of the *Kappan*, "Are Standards in U.S. Schools Really Lagging Behind Those in Other Countries?"[2] Husén's analysis of the large discrepancies between nations in the proportion of youngsters enrolled in high school served as a basis for the first section of my article.

Bradburn and his colleagues then argue that, while the researchers' aspirations have been only partially realized because of the many problems associated with these studies, publications of the International Association for the Evaluation of Educational Achievement (IEA) can teach us much about education in other countries that cannot be learned in any other way. That is true. Unfortunately, a lot of what we learn is inaccurate. We have publicized a set of rankings that are seriously biased because of methodological flaws. In fact, we don't know what the relative rankings of nations would be if it were possible to control for the inadequacies of sampling.

The fact is that the quality of our scientific output and the skills of our science and engineering majors are extremely high.

For example, consider some findings reported in my December article. The IEA mathematics assessment ranks Japan first in a multinational comparison of eighth-grade mathematics, with Hong Kong in the middle. By the 12th grade, when only 3% of Hong Kong's young people are taking mathematics (compared to 12% in Japan), Hong Kong comes in first and Japan second. The reality is that Hong Kong's *schools* are not dramatically better in the 12th grade than in the eighth; the changed rank is simply a matter of extreme student selectivity in Hong Kong.

Similarly, in eighth-grade comparisons

The fact is that we don't know with any degree of accuracy what the rankings would be if the comparisons themselves were sound.

Hungary ranks near the top on international tests. However, Hungary enrolls 50% of its students – more than any other country – in 12th-grade mathematics courses. Not surprisingly, by the 12th grade, Hungary scores among the bottom countries. Are Hungarian high schools that much worse than Hungarian middle schools? Or does the normal pattern – more students, lower scores – explain the results?

By contrast, England/Wales, where only 6% of the students take 12th-grade mathematics, ranks among the top countries in the 12th-grade comparisons – a significant improvement over its rank in the bottom half in most of the eighth-grade comparisons. Did the schools improve, or is it more likely that greater selectivity simply resulted in higher average test scores for those relatively few students who take mathematics in the 12th grade?

The fact is that we don't know with any degree of accuracy what the rankings would be if the comparisons themselves were sound. While the IEA reports contain useful information about curricula, that information can be gathered without conducting "test score competitions" that lead to inaccurate conclusions about the quality of educational practice.

Bradburn and his colleagues go on to present an algebraic equation to show that sampling error could not account for Korea's apparent superiority over the U.S. in comparisons made by the International Assessment of Educational Progress – a rather surprising rebuttal to my article, since I took no position one way or another about the relative rankings of Korea and the U.S. or of any two nations. Moreover, the use of an equation that relies on *hypothetical* numbers to estimate the accuracy of the relative rankings makes my point. Why must the numbers be hypothetical? We simply do not have the empirical data about which population groups were actually represented – or not represented – in the comparisons. Without these data, no equation can tell us what the rankings would have been if the sampling problems were eliminated.

In a related argument, the rejoinder notes that "sampling alone cannot come close to accounting for the superiority of other nations' performance." Certainly we all agree that many factors contribute to the rankings. My article, for example, cited differences in curricula and teaching strategies across countries, differences in the proportion of low-income students taking the test, the relative emphasis given to subjects in one country as opposed to another, and the representativeness of the items chosen to measure mastery of subject matter.

But it simply does not follow that, because sampling error might not "come close" to explaining the differences between countries, the quality of U.S. educational practice does! In short, the argument that the "superiority" of other nations in the rankings can be attributed to an inadequate U.S. education system is not supported by the studies. The fact is that comparisons of test scores tell us very little about the quality of education in any country.

My December article proposed a number of alternative criteria for assessing our accomplishments and deficiencies in science and engineering. For example, how productive is the U.S. in basic and applied research? What does the marketplace say about the research opportunities in our institutions of higher learning? Where are students from other parts of the world taking their advanced degrees in science and engineering? What are our accomplishments and failures in making major technological advances? Are we successful in turning our scientific advances into products that international consumers want to buy? Are science and engineering fields attracting high-achieving students? Is there a shortage of students or faculty members in science and engineering? Are we making prog-

"I hear that summer classes are really speeded up."

ress in attracting women and minorities to science and engineering? Does the teaching environment in our schools and colleges encourage students to select – and continue to study – science and mathematics? Does the educational experience give students who do not major in these fields a meaningful understanding of key scientific issues and methods? Do we provide the general student population with the skills needed to be competitive and productive in the workplace? Are we maintaining the technical expertise of our work force? Can there be any doubt that these measures of a nation's scientific achievements and problems are crucial – and certainly more relevant – than the outcomes of paper-and-pencil tests of noncomparable samples of students?

However, Bradburn and his colleagues question whether any single indicator, such as the number of scientific publications, can be "free from . . . incidental factors" that are not germane to the quality of schooling. I certainly did not suggest so. While no indicator by itself can tell the whole story, some are better than others, and a wider range of indicators gives a clearer sense of a nation's accomplishments and concerns. In any case, I am not proposing that we conduct a series of highly publicized "super bowls" to select winners and losers in a contest of scientific publication.

The rejoinder then questions whether substituting the kind of indicators that I suggest "may make us feel better as a nation, but . . . not help us very much in improving our education system for the future." (Will test score comparisons?) My purpose in suggesting a wide range of indicators that deal with current accomplishments and future trends is not to improve our mental health but to turn our attention to important policy matters and to provide insights into the areas that most need attention.

FINALLY, THE rejoinder argues that forthcoming studies will do a better job of addressing the problems than previous studies have. Perhaps. But the evidence suggests the contrary.

For nearly three decades, a great deal of expertise has been applied to the methodological problems of international assessment. And it is precisely because previous researchers understood the problems so well (but were unable to solve them) that I am not optimistic about the quality of future studies. Indeed, with the participation of many more countries, the methodological and practical problems can be expected to become even more troublesome.

Furthermore, it would clearly be inadvisable to undertake the "heroic" measures that would be needed even to begin to address the methodological problems. Would our children's education improve if we established rigid international controls on each nation's sampling design or attempted to locate out-of-school (or homeless) children and test them on science and mathematics? 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? Yet that is the primary rationale for conducting the studies in the first place.

> **Let's focus our attention on the difficult public policy issues to be addressed rather than on comparisons and rankings.**

Indeed, recent discussions of this matter suggest that my original article touched only the tip of the iceberg as far as the range of problems and the difficulty of reaching meaningful conclusions about the comparative quality of education are concerned. Consider, for example, the implications for international assessments of the following practices:

- In some countries, significant numbers of low-achieving schools – or schools in which the curriculum is considered to be inadequate – are excluded from the comparisons.
- In other countries, many students who are in industrial apprenticeship programs do not participate in the test comparisons.
- Several countries track students for all subjects in separate classrooms or separate schools as early as 10 years of age. We don't know which students are represented in the test comparisons, and therefore we can't determine the reasons for a particular level of test performance.
- Each nation's division of students by language, social class, ethnicity, race, religion, immigration status, region, public or private school, and academic or vocational school differs. We simply do not have the data to understand how all this plays out: who is – or is not – tested and what their educational experiences have been. We are not even able to describe clearly the various countries' education systems – let alone devise an appropriate sampling design that would enable us to look at the outcomes of those systems and "rank" them, or determine how much differences in the quality of education account for the test score differences.
- The problems are compounded in developing countries. Because of scarce resources, these countries typically have strongly elitist education systems that provide a high-quality education to relatively few, highly selected students. Indeed, a large proportion of students in developing countries have left school by the time the tests are administered.

It should not be surprising that our studies are flawed or that we have not yet developed a high-quality design for future studies. But even if we could, how would all these tests improve education for the children in any of the participating countries? There are a great many important questions related to the state of science and engineering education in the U.S. and to the matter of which students need to be better served. Let's get on with it and focus our attention on the difficult public policy issues to be addressed rather than on comparisons and rankings.

---

1. Iris C. Rotberg, "I Never Promised You First Place," *Phi Delta Kappan*, December 1990, pp. 296-303.
2. Torsten Husén, "Are Standards in U.S. Schools Really Lagging Behind Those in Other Countries?," *Phi Delta Kappan*, March 1983, pp. 455-61. K