

MYTHS IN INTERNATIONAL TEST SCORE COMPARISONS
AND WHERE THEY LEAD US

Princeton City School District, Cincinnati, Ohio
March 23, 1993

Thank you for inviting me this evening. I feel a little like a speaker discussing the Johnstown, Pennsylvania, flood--with Noah sitting in the audience. All of you have had a lot of experience in the areas I will discuss.

Since the 1960s, there have been a number of international test score comparisons of science and mathematics achievement, sponsored by different organizations and involving different countries. In my remarks today, I will summarize the shortcomings of these studies and then discuss related issues and questions that are often raised.

I have two main concerns about the international comparisons:

o First, international test score comparisons are flawed methodologically and do not reflect the quality of education in any country.

o Second, a reliance on a narrow criterion--answers on multiple choice tests--ignores far more important measures of our strengths and shortcomings in science education and leads us to recommend solutions that are irrelevant at best and often are counterproductive to resolving or even addressing our most important problems.

The basic problem in any comparisons of this type is that the more students who take the test, the lower will be the average score. That score has little to do with the quality of education. It simply reflects the fact that the students represented in the test comparisons are much more highly selected in some countries than in others.

o In the 1960s, high school attendance rates in the United States were substantially higher than those in most other countries. United States: close to 80 percent; European average: 20 percent.

o Major reversals of rankings between higher and lower grades: e.g., Hungary and England/Wales, Japan and Hong Kong, the former Soviet Union, Slovenia, and the United States.

o Analogous to SAT scores: ranking of states; letter from resident of affluent (and therefore high SAT) district that merged with low-income, low SAT district.

Again, the important point is that the more students who take the test, the lower will be the average score. That finding holds

whether the comparisons are among nations, among states, among school districts, or among schools within a district.

The second major factor contributing to the international test scores is the proportion of low-income children in the test-taking population. The United States has a relatively high proportion of low-income students compared to many other industrialized countries. And the numbers of low-income students have grown substantially in the past decade. We tend to hold the education system responsible for our broader societal problems.

Third, differences in curriculum emphases among nations also affect the international rankings--e.g., differences in timing of calculus courses.

Finally, there are differences in how the material is presented. Bette Bao Lord (the author of Spring Moon) puts it this way:

"As a fifth grader in Brooklyn's P.S. 8 . . . even before I had mastered fifty words of Brooklynese my teacher . . . began asking me for my opinion on every matter . . . I was flabbergasted by the fact that an adult--and not just any adult; on the contrary, my most honorable teacher--would solicit the opinion of a child--not just any child; on the contrary, an eight-year-old immigrant just off the boat. . . . And before long I came to realize that the merits of one's opinions were not the crucial point of the exercise. The crucial point was to air whatever opinions one had, and today I value this aspect of what we Americans delight in praising as our way of life perhaps more than any other."

The emphasis on classroom discussion that Bette Bao Lord describes, while highly desirable, is not necessarily reflected in higher scores on multiple choice tests of isolated pieces of information.

I am often asked whether with all our expertise in statistics and sampling design we can't simply improve the validity of the international comparisons. I don't believe we can. The problems are endemic to all of the studies since they began in the 1960s. The problems are not a matter of statistical expertise, but of the societal and educational diversity among countries. This diversity cannot be controlled for by any statistical design.

There are large differences among countries in which students take the test. For example:

- o Exclusion of 20 percent of the classes.
- o Enrollment in apprenticeship programs.
- o Tracking by age 11.
- o Exclusion of regions, language groups; in most recent ETS study, for example, Italy included only one province (Emilia-Romagna), the former Soviet Union (when it still was the Soviet Union) included

only Russian-speaking schools, Israel included only Hebrew-speaking schools.

o Highly specialized curriculum in some countries. Even Princess Diana (not then a princess, of course) did not continue in an academic program past the age of 16.

o Problems magnified by inclusion of broad range of developing countries, with highly elitist school systems (because of scarce resources) and substantial proportions of children who are out of school and therefore do not take the tests. The international comparisons are no more useful to a developing country than they are to the United States.

o The most recent ETS international comparison shows that no study, however well-intentioned, can address these problems. ETS went out of its way to point out the problems and advised in its press release against ranking the countries.

I am sometimes asked whether I believe we can learn something from other nations' education systems or teaching practices.

Of course we can. The challenge is to identify those practices that can realistically be transferred from one nation to another.

o However, in most cases, it would involve a basic restructuring of a nation's social, cultural, and political norms, including changes in the respective roles of national and local governments in education, the role of the teacher in society, teachers' salaries, comprehensive high schools, 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.

o Even when there is a public discussion within a country about making basic changes in education, the nation's social and cultural norms make it very difficult to accomplish. For example, in the United States, the current debate is about giving our elementary and high school students a more demanding curriculum, and then administering national tests; Japan would like its students to express their own views more readily; Taiwan would like its students to play more! A matter of culture, not the education system. In the area of international competitiveness, there are similar problems in trying to adapt industrial policies from Japan or Germany to the United States because government/industry links differ so fundamentally between the countries.

Another topic that is often raised is 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 do not relate to

weakness in science education or international test comparisons, but to business practices, government policies, and the realities of a global economy. Examples are exchange rates, the lack of incentives for industry to invest in long-term product development, 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--although that is beginning to change. These are far more important explanations of the status of U.S. competitiveness than are rankings on international test comparisons.

I would like to turn now to the issue of whether it matters if we focus on test score differences, or if we exaggerate the problems in the United States, when we all agree that our education system can be better than it is.

I believe it does matter.

First, 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. The fact is that our productivity in scientific research is extremely high--Nobel prizes, scientific publications, high quality scientists and engineers--and no shortages.

Indeed, there has been so much publicity about potential shortages of scientists and engineers in the past decade that many students chose careers based on inaccurate projections. Yet, we now know that many scientists and engineers are having difficulty finding jobs, e.g., there were more than 800 applications for one position in the physics department at Amherst College.

Second, the rhetoric detracts from our real problems--the large proportion of our children who live in poverty, the vast differences in educational resources between rich and poor schools, and the combination of rising costs of higher education, reductions in the real value of student financial aid, and decreasing state expenditures for higher education--and what these trends do to student motivation. My concern is that a focus on test scores deflects attention from what we can do to solve our real problems.

Consider, for example, the differences in educational resources between rich and poor schools.

- o Expenditures in some states are twice as high as in others.

- o Within states, expenditures in some districts are two to three times as high as in others: In Illinois, school districts spend between \$2400 and \$8300 per pupil. The 100 poorest districts in Texas spend an average of just under \$3000 per student; the 100 wealthiest districts, however, spend an average of about \$7200. In Mississippi, the range is between \$1300 and \$4000 per pupil. A

judge in a school finance case put it this way: "If money is inadequate to improve education, the residents of poor districts should at least have an equal opportunity to be disappointed by its failure."

o Large inequalities in education expenditures occur within school districts, as well as among districts and states. Some schools have half the funding of other schools in the same district. On average, those schools with high proportions of low-income and minority students receive less money. Examples: Rodriguez vs. Anton school finance litigation in Los Angeles; experience of principal who moved from West Side of New York to Harlem.

o School finance inequalities are not simply theoretical accounting matters. They have a direct effect on the expertise and education of teachers, on the teacher-student ratio, on the availability of counselors, on curriculum and instructional materials, on instructional methods--and, of course, on the physical environment of the school. And these inequalities have the greatest impact on precisely the students who are most neglected by the American education system--the students who will not attend college. Jonathan Kozol (in Savage Inequalities) has provided a picture of just how important money is by comparing what students in wealthy Chicago suburbs receive in school compared to students who attend inner-city schools:

The suburban New Trier High School students can take advantage of superior science laboratories, up-to-date technology, and music and art facilities. Latin and six other foreign languages are offered. Electives include the literature of Nobel prize winners, aeronautics, criminal justice, and computer languages. Average class size is 24 children . . . special education classes hold 15 . . . each student has access to a personal counselor free to work with only two dozen students. At the inner-city schools Kozol visited, in contrast, laboratories are makeshift or nonexistent, there are no music or art classes . . . no playground . . . compensatory education classes hold up to 40 students . . . and guidance counselors struggle with student loads of 420.

How do the children in inner-city schools perceive these differences? Kozol asked a student:

"If the governor announced that he was going to combine you with the kids from Cherry Hill--everybody goes to one school maybe for the ninth grade and the tenth grade, everybody to the other school for both their final years--what would you say?" "As soon as it was announced they'd start remodeling," Luis replies. "You'd see progress very fast. Parents of white children, with their money, they'd come in and say, 'we need this fixed. Out kids deserve it.' So they'd back us up, you see, and there'd be changes."

The point is that a preoccupation with test scores diverts our attention from what we can do to solve our real problems. It also leads us to recommend solutions that are marginal at best and may

be counterproductive. Recent proposals for national testing are an example. These proposals raise a number of concerns:

o First, if we implement the tests without reducing school finance inequalities, the tests are likely to do more harm than good. They may reduce graduation rates and subsequent employability and earnings, increase tracking, and generally screen out of the education system those students who already receive the lowest quality education. 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?

Harold Howe II, a former U.S. Commissioner of Education, describes the potential impact of national testing on students from low-income backgrounds, who have major problems to overcome both outside and inside their schools:

"Inside their schools, they are subjected to the effects of lower educational expenditures per student--larger classes, limited special services, decaying and inadequate facilities, 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."

o A second concern about national testing is that an emphasis on multiple choice standardized tests encourages the teaching of a narrow set of measurable skills that often have little to do with what we most value in education. Unfortunately, the schools with high proportions of low-income and minority students are more likely to emphasize rote learning than are more affluent schools. The low-income schools are under more pressure to raise standardized test scores.

According to one argument, testing can be improved by developing innovative new tests, often called "authentic tests," which would include performance assessments, essay exams, and portfolio assessments. These tests do not yet exist for purposes of national testing. They will be extremely expensive to develop and administer: Even after many years of development, it is questionable whether they would produce nationally reliable and comparable data. The most likely outcome is more funding for researchers and policy analysts--like myself--with little benefit to schools.

o Moreover, regardless of the quality of the tests, test score differences from year to year, or from one district or school to another, tell us little about the quality of the educational program. The tests tell us only what we already know--the effects of inadequate resources and poverty on the learning experience. Examples: Chapter 1 program improvement requirements and effective schools research.

I would like to conclude by noting that the environment for

education today is far more challenging than the environment even a decade ago when my daughters attended the public schools. The numbers of poor children have increased substantially. In recent years, several "silver bullet" proposals--national testing, which we discussed, national standards, vouchers, "restructuring"--have been put forward as the reforms needed to strengthen the nation's education system. The proposals do not begin to address either the severe problems of poverty in our inner-city and rural schools or the serious underfunding of these schools. These problems deserve a lot more attention than international--or domestic--test score comparisons and rankings.

Thank you.