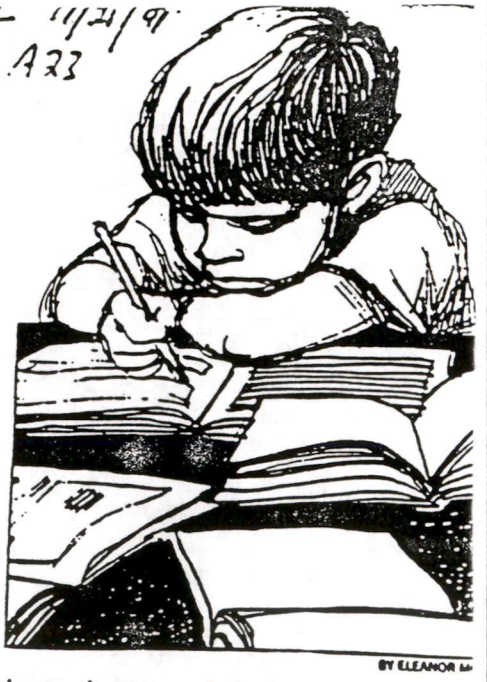
Iris C. Rotberg

Wash Post 11/21/91
p. A23

What Test Scores Don't Measure



BY ELEANOR M

In a recent op-ed piece on the quality of U.S. education ["U.S. Schools: The Bad News Is Right," Nov. 17] Diane Ravitch suggests that "the greatest obstacle to those who hope to reform American education is complacency." I agree. As a nation, we are too complacent about the large proportion of our students who are in poverty, about the vast disparities in educational expenditures between rich and poor school districts, about the rising costs of higher education and what it does to student motivation. But these serious problems will not be addressed by international test comparisons that are seriously flawed and, in fact, irrelevant.

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

Taking Exception

nations. Yet, after almost three decades of apparent failures on international tests, we have maintained a level of productivity in science and engineering that is overwhelming. The fact is that international comparisons of test scores are highly misleading indicators of the quality of a nation's education system or the expertise of its future scientists and engineers.

The rankings of nations in international test comparisons are meaningless because it is virtually impossible to control for the major societal differences among nations. For example, attendance rates in the final years of high school are much higher in the United States than in most other countries. Indeed, the first international assessments compared the average score of more than 75 percent of the age group in the United States with the average score of the top 9 percent of the students in West Germany, the top 13 percent in the Netherlands and the top 45 percent in Sweden. 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 in any country.

Consider, for example, the results of a recent assessment of mathematics students in Hungary and England. Hungary ranks near the top in the eighth-grade comparison. Not surprisingly, by the 12th 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 12th grades, or is it simply a matter of more students, 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 12th grade, when only a highly select group of students there takes the test. Similarly, in eighth-grade comparisons, Japan ranks first, with Hong Kong in the middle of the rankings. By the 12th grade, when only 3 percent of Hong Kong's young people are taking mathematics (compared with 12 percent in Japan), Hong Kong comes in first and Japan second.

When a country's rank can change so dramatically between the eighth and 12th grades, it simply shows that the test comparisons are meaningless.

It is not just a matter of student attendance rates. For example, it has been observed that in a recent mathematics assessment of 13-year-olds in six countries, 99 percent of the age group attended school. What isn't stated is that the samples of children actually tested were not representative of the entire country. Thus, the entire United States was compared with only selected Canadian provinces. Only the largest of several language groups in Spain participated in the comparisons. The Inner London Educational Authority chose not to participate in the assessment.

Moreover, some countries exclude from the testing significant numbers of low-achieving schools and schools in which the curriculum is considered inadequate. Several use the track system, separating students according to ability as early as 11 years of age. We do not know which students are represented in the test comparisons.

In other countries, students take courses almost exclusively in their fields of specialization after age 16. Thus high school students who are tested in science and mathematics have studied essentially *only* science and mathematics from age 16 on.

The problems are magnified enormously by the inclusion of a much greater range of countries in forthcoming studies. China illustrates the problem. Like many other developing countries with scarce resources, China has a highly elitist education system that provides advanced mathematics and science instruction to very few selected students. The majority of Chinese young people have either left school by the age the test is administered or have never studied the material covered by the assessment and are unlikely to be represented in the sample taking the test. A comparative assessment, therefore, is meaningless if the test is given only in selected schools.

Differences in the incidence of poverty among students taking the test also affect the rankings. Countries with substantial proportions of low-income students taking the test 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.

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

But there is a more fundamental issue. Even if the test results accurately portrayed the relative "rankings" of participating countries, we are still left with the matter of whether test scores are a useful measure of those things that are most important.

The fact is that the quality of our scientific output and the skills of our science and engineering majors are extremely high. While our success in turning research into marketable products is questionable, our international competitiveness relates less to weakness in science education or international test comparisons than to far more subtle factors: the lack of incentives for industry to invest in long-term product development, financial incentives that lead to offshore manufacturing, licensing practices and the emphasis placed on military at the expense of civilian research.

These issues will not be addressed by yet another round of international tests. Nor will test comparisons provide a better education for low-income students who attend schools with inadequate resources. These are the real problems we should not be complacent about. Let's focus our attention on the difficult public policy issues to be addressed rather than on spurious comparisons and rankings.

The writer has directed research programs at the National Science Foundation and the National Institute of Education.