Comment on "Too Little Too Late: American High Schools in an International Context" by William H. Schmidt in *Brookings Papers on Education Policy: 2003*, Diane Ravitch, Editor.

Comment by Iris C. Rotberg

The validity of research depends to a large extent on the validity of the assumptions on which it is based. Some of the assumptions in William H. Schmidt's paper have been tested and found to be accurate. Others have been tested and found to be questionable. For others, little evidence exists one way or the other. Examining the assumptions underlying the paper is useful, therefore, as a basis for interpreting the findings and assessing their implications for public policy. These assumptions also have general relevance because they appear frequently in the education research literature.

I begin with basic assumptions that have been tested and found to be valid. First, students who have studied the material covered by a test will get higher test scores than those who have not. It is difficult to do well on a calculus test if you have never studied calculus. Second, the paper assumes a selection bias—that is, some students are advised, inappropriately, to take less demanding courses. That advice, in turn, reduces the students' potential for high academic achievement and closes options they otherwise might have had. The important point is that schools have a responsibility to offer each child the strongest possible educational experience.

From these well-tested assumptions, the paper moves to a set of assumptions for which there is less support. First, the paper assumes that the Third International Mathematics and Science Study (TIMSS) provides the information needed to draw conclusions about the quality of education in participating countries. The main problem is that Schmidt uses TIMSS data as the basis for his analysis.[35] In my view, TIMSS is so flawed that implications cannot be drawn about the quality of education in any of the participating countries.[36] Rankings of countries in TIMSS are based on simple comparisons of test scores in the final year of secondary school, without any controls for the large differences between participating countries on a wide range of variables. Therefore, it is impossible to learn from the study how variables such as the following affected student selectivity and, in turn, test score rankings: participation and exclusion rates of both schools and students being tested; percent of the age cohort who dropped out of school and therefore did not take the test; percent of students taking advanced assessments; average age and grade of students taking the test; special concentrated programs for different students; practices with respect to the inclusion or exclusion of low-achieving students, language minority students, students with disabilities, apprenticeship programs, and entire regions of the country in the test comparisons; the mix of public and private schools, comprehensive and specialized schools, and academic and vocational schools; tracking and coaching practices; family socioeconomic status (SES); and the consistency between the education program and the test.

Each of these variables can be expected to play a significant role in the extent to which the students taking the test represent a highly select group, not the general student population. The TIMSS study did not conduct a multivariate analysis to provide information about the contribution of each variable to the test score rankings. The variables are so confounded that how any of them, individually or in combination, affected the test scores cannot be determined. The use of these data, therefore, does not contribute either to research knowledge or to informed public policy.

The difficulty of unraveling the TIMSS findings is illustrated by tables 1–3, which show the wide differences between countries on several of the major variables. The tables, which are adapted from data presented in the TIMSS report, also explicitly show that few of the participating countries

Table 1. TIMSS Scores on Assessments of Mathematics and Science General Knowledge, Advanced Mathematics, and Physics

Nation	Average mathematics score	Average science score	Average advanced mathematics score	Average physics score
Australia	522	527	525	518
Austria	518	520	436	435
Canada	519	532	509	485
Cyprus	446	448	518	494
Czech Republic	466	487	469	451
Denmark	547	509	522	534
France	523	487	557	466
Germany	495	497	465	522
Greece	—	—	513	486
Hungary	483	471	—	—
Iceland	534	549	—	—
Italy	476	475	474	—
Latvia	—	—	—	488
Lithuania	469	461	516	—
Netherlands	560	558	—	—
New Zealand	522	529	—	—
Norway	528	544	—	581
Russian Federation	471	481	542	545
Slovenia	512	517	475	523
South Africa	356	349	—	—
Sweden	552	559	512	573
Switzerland	540	523	533	488
United States	461	480	442	423
International average	500	500	501	501

Source: Adapted from data presented in Department of Education, National Center for Education Statistics, *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*, NCES 98–049 (Government Printing Office, 1998), figures 1, 5, 9, and 16.

Note: TIMSS = Third International Mathematics and Science Study. Nations not meeting international sampling and other guidelines are in italics. Canada and France met the guidelines for the advanced mathematics assessment; France and Norway, for the physics assessment. A dash (—) indicates that the nation did not participate in that assessment.

met the international sampling and other guidelines set forth by the TIMSS researchers.

Schmidt uses TIMSS data as the main basis for his findings and recommendations with respect to curriculum. Unfortunately, the study provides no guidance about curriculum or any other component of school systems. It serves, instead, primarily as a Rorschach test that reflects previously held views about U.S. schools.

Second, the paper assumes that the performance of U.S. schools declines from the fourth grade on. I do not know if the performance of U.S. schools

Table 2. TIMSS Sampling Data Participation and Exclusion Rates
Percent

Nation	*Mathematics and science general knowledge assessment*		*Advanced mathematics assessment*		*Physics assessment*	
	Combined participation rates	*Exclusion rates*	*Combined participation rates*	*Exclusion rates*	*Combined participation rates*	*Exclusion rates*
Australia	52	6	55	No data	54	No data
Austria	73	18	81	18	81	18
Canada	68	9	77	No data	73	No data
Cyprus	98	22	96	22	96	22
Czech Republic	92	No data	92	No data	92	No data
Denmark	49	2	49	No data	47	No data
France	69	1	77	No data	77	No data
Germany	80	11	78	11	82	11
Greece	—	—	87	No data	87	No data
Hungary	98	0	—	—	—	—
Iceland	74	0	—	—	—	—
Italy	62	30	68	30	—	—
Latvia	—	—	—	—	77	50
Lithuania	85	16	92	16	—	—
Netherlands	49	22	—	—	—	—
New Zealand	81	0	—	—	—	—
Norway	71	4	—	—	83	No data
Russian Federation	90	43	96	43	95	43
Slovenia	42	6	42	No data	43	No data
South Africa	65	0	—	—	—	—
Sweden	82	0	89	No data	89	No data
Switzerland	85	3	87	No data	87	No data
United States	64	4	67	No data	68	No data

Source: Adapted from data presented in Department of Education, National Center for Education Statistics, *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*, NCES 98–049 (Government Printing Office, 1998), tables A1.1, A1.2, and A1.3.

Note: TIMSS = Third International Mathematics and Science Study. The sampling plan established the following protocol for selecting schools and students to participate in the testing: (1) the sample was to be representative of at least 90 percent of students in the total population eligible for the study (that is, exclusion rates should be no greater than 10 percent); (2) the school participation rate without the use of replacement schools should be at least 50 percent; and (3) the combined participation rate (computed by multiplying the school and student rates after replacements) should be at least 75 percent or school and student participation rates each should be 85 percent. Nations not meeting international sampling and other guidelines are in italics. Canada and France met the guidelines for the advanced mathematics assessment; France and Norway, for the physics assessment. A dash (—) indicates that the nation did not participate in that assessment. "No data" indicates that the nation participated, but the TIMSS report did not provide data on combined participation rates or exclusion rates.

Table 3. TIMSS Data on School Completion Rates, Percent of Age Cohort Taking Advanced Subjects, Age and Grade of Participating Students, and Differentiation in Programs

Nation	*Percent of twenty-five- to thirty-four-year-olds completing secondary education*	*Percent taking advanced assessments as proportion of age cohort*		*Average age of participating students*	*Grades of participating students*	*Extensive differentiation in programs for students with differing abilities or interests*
		Advanced mathematics	*Physics*			
Australia	57	16	13	17.7	12	No
Austria	81	33	33	19.1	10–14	Yes
Canada	84	16	14	18.6	12–14	No
Cyprus	No data	9	9	17.7	12	Yes
Czech Republic	91	11	11	17.8	10–13	Yes
Denmark	69	21	3	19.1	12	Yes
France	86	20	20	18.8	11–13	Yes
Germany	89	26	8	19.5	12–13	Yes
Greece	No data	10	10	—	—	No data
Hungary	No data	—	—	17.5	10, 12	Yes
Iceland	No data	—	—	21.2	12–14	Yes
Italy	49	14	—	18.7	11–13	Yes
Latvia	No data	—	3	—	—	No data
Lithuania	No data	3	—	18.1	12	Yes
Netherlands	70	—	—	18.5	11–12	Yes
New Zealand	64	—	—	17.6	11–12	No
Norway	88	—	8	19.5	12	Yes
Russian Federation	No data	2	2	16.9	11	Yes
Slovenia	No data	75	39	18.8	11–12	Yes
South Africa	No data	—	—	20.1	12	No
Sweden	88	16	16	18.9	11–12	Yes
Switzerland	88	14	14	19.8	11–13	Yes
United States	87	14	14	18.1	12	No
International average	78	19	14	18.7	—	

Source: Adapted from data presented in Department of Education, National Center for Education Statistics, *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*, NCES 98–049 (Government Printing Office, 1998), tables A5.7, A5.12, A5.13, and A5.14.

Note: TIMSS = Third International Mathematics and Science Study. Nations not meeting international sampling and other guidelines are in italics. Canada and France met the guidelines for the advanced mathematics assessment; France and Norway, for the physics assessment. A dash (—) indicates that the nation did not participate in that assessment. "No data" indicates that the nation participated, but the TIMSS report did not provide data on school completion rates or program differentiation.

declines between elementary school and high school, but the test score rankings do not inform the issue. Indeed, the TIMSS report presents data showing that the average age of the students taking the test in each country influences its relative performance between eighth grade and the final year of secondary school. In the general mathematics assessment, five countries ranked higher in the final year than in eighth grade, six (including the United States) ranked lower, and nine maintained their position. However, the TIMSS report points out that the countries that declined had the smallest average age gap between the two grades (3.5 years), whereas those that gained had the largest age gap (5.4 years). In short, the findings are an artifact of the research design: Students in some countries were older and had more years of schooling.

Third, the paper states that "the causative factors" [for the test score rankings] "most likely lie with the educational system and not with the nature of the students or associated demographics." That assumption is not supported by the current study or by the significant body of research that shows a strong relationship between low student achievement and poverty. TIMSS did not collect the data needed to quantify the relationship between student poverty and test scores in the international comparisons. However, Schmidt—after making the statement quoted above—goes on to demonstrate a significant correlation between SES and student achievement based on U.S. data. The countries participating in TIMSS differ substantially in rates of absolute and relative poverty, both of which can be expected to play a major role in the test score rankings. Furthermore, the United States ranks high in relative poverty.[37] Looking at the research literature more broadly, it is difficult to identify any other correlation in educational research that has been so consistent and pronounced as between poverty and low educational achievement. For example, the test score rankings of states on the National Assessment of Educational Progress (NAEP), the rankings of districts within states, or the rankings of schools within districts clearly show that correlation.

That does not mean children from low-income families cannot excel in school. Many overcome the odds and achieve at a high level. But the odds are not on their side, both because of poverty and because the schools they attend receive the fewest resources. Constance Clayton, the former superintendent of the Philadelphia public schools, put it this way: "We must face every day the realities of the unequal hand dealt to our children and to our schools."[38] Curriculum changes will not address these basic educational problems.

Fourth, the paper assumes that test score rankings are valid predictors of both a nation's productivity in science and technology and its economic strength. Student test scores have been used for almost fifty years to explain a variety of perceived crises. In the 1950s and 1960s, following the launch of *Sputnik*, U.S. analysts were concerned that the country might not be able to compete with the Soviet Union in technology. Later, shortages of scientists and engineers were predicted. That crisis was followed by a concern about competing economically with Japan. The rhetoric linked each of the perceived problems to international test score rankings. Yet, the United States has maintained a high level of productivity in science and technology, as measured by basic research, technological advances, and product development. Moreover, the economy has generally been strong, with one of the lowest unemployment rates in the world. Clearly, some U.S. schools have real problems. However, these problems will not be solved by making tenuous links between test score rankings on TIMSS and what the society perceives as its current crisis. Little evidence exists that the rankings of industrialized nations on international test score comparisons predict either a nation's productivity in science and technology or its economic strength.

Fifth, the paper states that "twenty percent of [the] five million [workers in the U.S. information technology sector] were foreign-educated and came to the United States specifically to fill an unmet need in this sector because of inadequacies in basic mathematics and science in the U.S. education system." While high school preparation in mathematics and science always leaves room for improvement, there is little evidence to draw a connection between staffing patterns in technological industries and the quality of curricula in U.S. schools. Do the data support a cause and effect relationship between U.S. schools and the participation of workers who were educated abroad? Is it possible, for example, that shortages of U.S. workers are caused by the fact that U.S. mathematicians prefer to become investment bankers or technology entrepreneurs, where the financial rewards are greater? Or, perhaps, the industries that report shortages are unable to find enough qualified U.S. residents at the salaries they choose to pay. Existing data do not provide evidence to choose among these, or other possible, interpretations.

Finally, the paper assumes that improving curricula and increasing and revising course requirements will address the most important educational problems. Schmidt hypothesizes that "much of the poor performance of the United States may be attributed to a poorly constructed curriculum that is not coherent from a disciplinary point of view and not intellectually rigorous from an international point of view." Perhaps, but the major educational problems in the United States occur in communities with high poverty rates and inadequate resources for education. It would be reassuring to believe that the most difficult educational problems could be addressed simply by revising curriculum, with little attention to the underlying causes of the problems. However, the evidence suggests that the problems are much deeper. They stem from poverty and the fact that the nation devotes the fewest educational resources to the students with the fewest resources. Quick fixes—whether more course requirements or more tests—will not address the basic issues. My concern is that, by focusing on what, at best, are marginal solutions, real problems are ignored.

Notes

35. Department of Education, National Center for Education Statistics, *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context*, NCES 98–049 (Government Printing Office, 1998).

36. The analysis of the Third International Mathematics and Science Study is based on Iris C. Rotberg, "Interpretation of International Test Score Comparisons," *Science*, May 15, 1998, pp. 1030–31. Tables 1–3 appeared at *Science* Online in connection with that article (www.sciencemag.org/feature/data/981368.shl [May 15, 1998]).

37. See, for example, McKinley L. Blackburn, *Comparing Poverty: The United States and Other Industrialized Nations* (Washington: American Enterprise Institute for Public Policy Research, 1997).

38. Iris C. Rotberg and James J. Harvey, *Federal Policy Options for Improving the Education of Low-Income Students*, vol. I: *Findings and Recommendations* (Washington: RAND, 1993), p. IV.