10-2013

Is the United States Winning or Losing the International Horse Race in Academic Achievement? Neither—It is Running With Other Western G7 Nations

Erling E. Boe
University of Pennsylvania, boe@gse.upenn.edu

Sujie Shin
University of Pennsylvania

Is the United States Winning or Losing the International Horse Race in Academic Achievement? Neither—It is Running With Other Western G7 Nations

Abstract
It is widely believed and lamented that students from the US perform poorly in international comparisons of academic achievement. Such perceptions have led to grave concerns about the future economic competitiveness of the US internationally. These concerns have been based on a generation of older international surveys on mathematics and science achievement. Fortunately, a recent generation of high quality international achievement surveys has been completed since 1990 on a wider array of subjects—reading, mathematics, science, and civics. Accordingly, the purpose of this report is to assemble and organize the results of all major international achievement surveys re-ported since 1990 in order to determine how well US students have performed in comparison with their peers from 21 other industrialized nations.

Upon aggregating the standing of US achievement scores across subject matters and grade levels, the results indicated that US students score somewhat higher than their peers in other industrialized nations, with only 24% of national scores being significantly higher than the US and 35% being significantly lower. Therefore, US students generally perform above average in international comparisons instead of poorly. The exception was mathematics, a subject in which US students score somewhat below average. It was also found that US students performed above aver-age at the elementary grade level, and average at the middle and secondary levels.

More detailed comparisons of achievement scores were made with the major economic competitors of the US—the G7 nations. At the middle and secondary grade levels (the levels at which the US is least competitive), US scores are comparable to those of other Western G7 nations in reading, mathematics, and science, and considerably higher in civics. Scores of Japanese students in reading are comparable to Western G7 nations and the US, but much higher in mathematics and science. With respect to academic achievement, the US is quite comparable to other major Western nations, whereas the Western G7 nations consistently trail Japan in mathematics and science.

Because of the well-known achievement gap in the US between White and minority students, scores were further analyzed by race/ethnicity (White, Black, and Hispanic). US achievement scores for the majority White students were consistently greater than those of the other five Western G7 nations, even though these nations were pre-dominantly White. By comparison, the scores for US Black and Hispanic students were very low and well below other scores. This is compelling evidence that the low scores of two minority groups were major factors in reducing the comparative standing of the US in international achievement surveys. That is, if these minority students per-formed at the level of US White students, the US would lead all G7 nations (including Japan) in reading and would lead Western G7 nations in mathematics and science, while still trailing Japan in mathematics.

We conclude that US students have generally performed above average in comparisons with industrialized nations instead of poorly as widely perceived. The misconception of poor US performance may be due to several reasons—inadequate information, unreasonable expectations that the US should be first-in-the-world, biased report-ing, and/or misleading comparisons of the US (a large multi-state nation) with small and homogeneous nations such as Finland and Ireland. In this respect, we compared TIMSS-linked science scores

of 40 US states with TIMSS science scores of 22 European nations (eighth grade level). The mean and variability of US science scores was very similar to that of scores from the European nations.

Disciplines
Education

IS THE UNITED STATES WINNING OR LOSING THE INTERNATIONAL HORSE RACE IN ACADEMIC ACHIEVEMENT? NEITHER—IT IS RUNNING WITH OTHER WESTERN G7 NATIONS[1]

Erling E. Boe, Ph.D.
Professor of Education

And

Sujie Shin, B.A.
Research Assistant

Research Report No. 2004TIMSS1

Center for Research and Evaluation in Social Policy
Graduate School of Education
University of Pennsylvania
Philadelphia, PA 19104

October 13, 2004

[1]Support for this research was provided by a grant (Award Number R215U980021) from the Fund for Improvement in Education (CFDA Number: 84-215U), Office of Reform Assistance and Dissemination, Office of Educational Research and Improvement, U.S. Department of Education; by a grant (Grant Number REC-9815112) from the Research on Education, Policy, and Practice Program (NSF 96-138), Division of Research, Evaluation, and Communication, Directorate for Education and Human Resources, National Science Foundation; and by the Center for Research and Evaluation in Social Policy, the Graduate School of Education of the University of Pennsylvania.

ABSTRACT

It is widely believed and lamented that students from the US perform poorly in international comparisons of academic achievement. Such perceptions have led to grave concerns about the future economic competitiveness of the US internationally. These concerns have been based on a generation of older international surveys on mathematics and science achievement. Fortunately, a recent generation of high quality international achievement surveys has been completed since 1990 on a wider array of subjects—reading, mathematics, science, and civics. Accordingly, the purpose of this report is to assemble and organize the results of all major international achievement surveys reported since 1990 in order to determine how well US students have performed in comparison with their peers from 21 other industrialized nations.

Upon aggregating the standing of US achievement scores across subject matters and grade levels, the results indicated that US students score somewhat higher than their peers in other industrialized nations, with only 24% of national scores being significantly higher than the US and 35% being significantly lower. Therefore, US students generally perform above average in international comparisons instead of poorly. The exception was mathematics, a subject in which US students score somewhat below average. It was also found that US students performed above average at the elementary grade level, and average at the middle and secondary levels.

More detailed comparisons of achievement scores were made with the major economic competitors of the US—the G7 nations. At the middle and secondary grade levels (the levels at which the US is least competitive), US scores are comparable to those of other *Western* G7 nations in reading, mathematics, and science, and considerably higher in civics. Scores of Japanese students in reading are comparable to Western G7 nations and the US, but much higher in mathematics and science. With respect to academic achievement, the US is quite comparable to other major Western nations, whereas the Western G7 nations consistently trail Japan in mathematics and science.

Because of the well-known achievement gap in the US between White and minority students, scores were further analyzed by race/ethnicity (White, Black, and Hispanic). US achievement scores for the majority White students were consistently greater than those of the other five Western G7 nations, even though these nations were predominantly White. By comparison, the scores for US Black and Hispanic students were very low and well below other scores. This is compelling evidence that the low scores of two minority groups were major factors in reducing the comparative standing of the US in international achievement surveys. That is, if these minority students performed at the level of US White students, the US would lead all G7 nations (including Japan) in reading and would lead Western G7 nations in mathematics and science, while still trailing Japan in mathematics.

We conclude that US students have generally performed *above average* in comparisons with industrialized nations instead of poorly as widely perceived. The misconception of poor US performance may be due to several reasons—inadequate information, unreasonable expectations that the US should be first-in-the-world, biased reporting, and/or misleading comparisons of the US (a large multi-state nation) with small and homogeneous nations such as Finland and Ireland. In this respect, we compared TIMSS-linked science scores of 40 US states with TIMSS science scores of 22 European nations (eighth grade level). The mean and variability of US science scores was very similar to that of scores from the European nations.

TABLE OF CONTENTS

Page

TABLE

FIGURES

Introduction

It is widely believed and lamented that students from the US perform poorly in international comparisons of academic achievement. For example, Silver (1998) stated that US students achieved *poorly* in mathematics at grades 7 and 8 in the Third International Mathematics and Science Study (TIMSS)[1], and that this indicated "a pervasive and intolerable mediocrity in mathematics teaching" (p. 1). Likewise from TIMSS, the Institute of Education Sciences (IES), USDE, referred to the *poor* performance in mathematics of US students in the middle grades, and attributed this to the ineffectiveness of mathematics education (IES, 2003). Such perceptions have led to grave concerns about the future economic competitiveness of the US internationally. For example, the Director of the National Science Foundation stated that the position of the US in the world economy depends critically upon the students achieving at high levels in mathematics and science (Colwell, 2000).

Given the concern about the US position in international economic competitiveness, the focus of US comparisons has narrowed to industrialized nations in particular. Citing the results of a generation of international surveys in mathematics and science achievement prior to 1992, Jaeger (1994) stated that US students "*typically* [italics added] scored well below students of similar age and school grade in the industrialized nations said to provide significant economic challenge to the United States" (p. 23).

Accordingly, it is often said that US students perform poorly in comparison with students from "many" industrialized nations. This indeed is true. One can pick a particular survey (e.g., TIMSS 1995), subject matter (e.g., mathematics), and grade level (e.g., grade 8) and find "many" industrialized nations that scored significantly[2] higher than the U.S. (e.g., France, Japan, and Switzerland). Yet it is also true that US students perform *highly* in comparison with students

[1] The original TIMSS (conducted in 1995) is hereafter referred to as TIMSS 1995. In addition, a TIMSS-Repeat was conducted in 1999. It is referred to as TIMSS 1999.

[2] When the word "significant" is used in this paper, it refers to a statistically significant difference between the achievement scores of two nations as found in reports issued by primary sources for the six international education surveys included here.

from "many" industrialized nations. For example, the US actually scored significantly higher than many industrialized nations (e.g., France, Germany, and Switzerland) in the Reading Literacy Study (RLS) at grade 4. Thus, depending on one's interest or agenda, one can pick a particular survey finding to support almost any vague conclusion about how the US stands in the international horse race.

Jaeger's (1994) conclusion about the US relative standing among nations is more credible than vague generalizations about "many" nations scoring higher than the US. This is because his conclusion was based on the results of several comparative surveys of mathematics and science achievement. Since his review, a new generation of international achievement surveys has been completed. Furthermore, these international surveys of achievement have not been limited to mathematics and science, but also include surveys of reading and civics achievement. These and other recent educational surveys are listed in Table 1, along with which of 22 industrialized nations participated in each survey at each grade level.[3] It is worth knowing how the US performed in these more recent surveys.

The common perception of poor performance in international comparisons by US students is typically attributed to ineffectiveness of American public education. Educators and policy makers of widely different perspectives embrace this inference because it creates enormous pressure for change. It is useful to those who are devoted to reforming public education in a wide variety of ways, as well as useful to those who are devoted to diminishing public education through various strategies leading to increased privatization such as through vouchers. With so much at stake, it is important to know just how well (or poorly) US students have performed in recent international achievement comparisons. Therefore, the purpose of this report is to assemble and organize the results of all major international achievement surveys reported since1990[4] to determine how well US students have performed in comparison with their peers from other indus-

[3] These 22 nations were classified as industrialized by the International Monetary Fund Statistics Department (2002) and as "High Income OECD Members" by the World Bank (2003). One country (Luxembourg) was also classified as industrialized by both sources, but did not participate in the international education surveys listed in Table 1 except for the PISA survey.

[4] Specifically the surveys listed in Table 1, all of which have been reported since Jaeger's (1994) review of US performance in earlier surveys. All have been conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), except for the Program for International Student Assessment (PISA) conducted under the auspices and the Organization for Economic Co-Operation and Development (OECD).

Table 1. Twenty-Two Industrialized Nations: Participation in International Education Surveys by Grade Level from 1991 through 2001

Industrialized Nation[a]	International Education Surveys								
	RLS[b]		PIRLS[b]	TIMSS 95[b]			TIMSS 99[b]	PISA[b]	CES[b]
	4th	9th	4th	4th	8th	FYS[c]	8th	10th	9th
Australia				x	x	x	x	x	x
Austria				x	x	x		x	
Belgium (Fr)	x	x			x			x	x
Canada (G7)				x	x	x	x	x	
Denmark	x	x			x	x		x	x
England[d] (G7)			x	x	x		x	x	x
Finland	x	x					x	x	x
France (G7)	x	x	x		x	x		x	
Germany (G7)	x	x	x		x	x		x	x
Greece	x	x	x	x	x			x	x
Iceland	x	x	x	x	x	x		x	
Ireland	x	x		x	x			x	
Italy (G7)	x	x	x			x	x	x	x
Japan (G7)				x	x		x	x	
Netherlands	x	x	x	x	x	x	x		
New Zealand	x	x	x	x	x	x	x	x	
Norway	x	x	x	x	x	x		x	x
Portugal	x	x		x	x			x	x
Spain	x	x			x			x	
Sweden	x	x	x		x	x		x	x
Switzerland	x	x			x	x		x	x
U.S. (G7)	x	x	x	x	x	x	x	x	x
N	17	17	11	13	20	14	9	21	13

[a]Industrialized nations as defined by the International Monetary Fund (2002) and the World Bank (2003).

[b]International education surveys and grade levels studied: Reading Literacy Study (RLS 1991) (USDE, NCES, 1996), Progress in International Reading Literacy Study (PIRLS 2001) (Mullis, et al., 2003), Third International Math and Science Study (TIMSS 1995) (Beaton, Mullis, et al., 1996), Third International Math and Science Study-Repeat (TIMSS 1999) (Mullis, Martin, Gonzalez, Gregory, Garden, et al., 2000), Program for International Student Assessment (PISA 2000) (OECD, 2001), Civic Education Study (CES 1999) (Torney-Purta, et al., 2001).

[c]FYS stands for "the final year of secondary school" (not grade 12).

[d]The United Kingdom participated in PISA.

trialized nations. More particularly, achievement data were analyzed to address the following questions:

- How do US achievement scores overall compare with scores from other industrialized nations?
- How does US performance compare with other industrialized nations by subject matter?
- How does US performance compare with other industrialized nations by grade level?
- How does US performance compare with its major economic competitors—the G7 nations?
- How does the performance of US racial/ethnic groups compare with other G7 nations?

General Method

National mean achievement scores were obtained from primary sources from 22 industrialized nations participating in six surveys as shown in Table 1. These sources also provided information on the statistical significance of the difference between the US mean achievement scores and the mean achievement score for each industrialized nation participating in a survey. These data are reproduced here in Tables A-1 through A-7 of Appendix A (pp.26-35) as originally reported in the primary sources identified in a footnote to each of these tables.

We aggregated the achievement data contained in Tables A-1 through A-7 to address the questions posed for this report. Procedures used for aggregation are described in Appendix B (pp 36-40). The outcomes of this process constitute the basic results reported here. All cross-national comparisons are based on national mean comprehensive scores (e.g., full-scale scores in mathematics) rather than on subscale scores (such as in algebra).

US Achievement by Subject Matter

The perception that the US performs poorly in international achievement comparisons is typically associated with mathematics. This perception may not be accurate, and ignores the US performance in other subjects.

Figure 1 shows the comparison of US achievement with other industrialized nations, aggregated across four grade levels, six surveys, and four subjects as listed in Table 1. In mathematics, about half the nations (53%) scored significantly above the US, while 32% of the nations scored at an equivalent level to the US. The remaining 15% of nations scored significantly below the

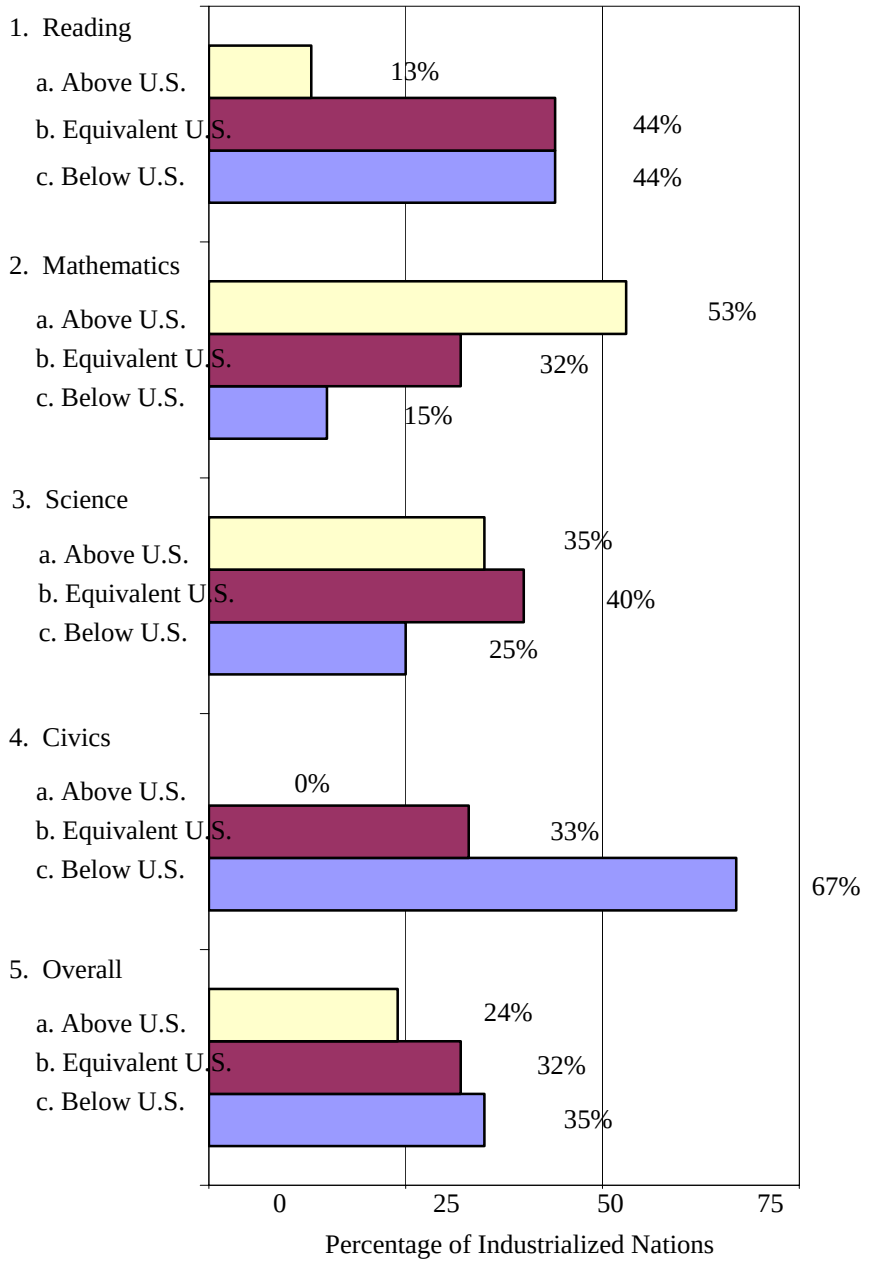

Figure 1. Percentage of 21 industrialized nations with national mean full-scale test scores significantly higher than, not significantly different than, and significantly lower than the comparable U.S. scores by subject matter across the 4th, 8th, 9th, and 10th grades, and during the Final Year of Secondary School. Sources: Five international education surveys sponsored by IEA and one by OECD.

US. Viewed this way, the aggregated results make clear that the US did not perform "poorly." Instead, the US mathematics score was somewhat below average. To say that the US performed poorly, one would expect at least 75% of industrialized nations to have scored significantly higher. Nonetheless, efforts to improve mathematics instruction and learning must continue, and perhaps be intensified, because only 26% of US students performed at the proficient (or above) level in mathematics as measured by the National Assessment of Educational Progress (NAEP) in the year 2000 (NCES, NAEP, 2000).

In contrast with mathematics, the US was much more competitive with other industrialized nations in three other subjects. As seen in Figure 1, the US performed at a high level in civics (no nation scored significantly higher in civics) and in reading (only 13% of nations scored significantly higher in reading), and was about average in science (with 35% of nations scoring significantly higher and 25% lower in science). A case can be made that learning *each* of the four subjects is a vital objective of public schools. Certainly reading is as basic as mathematics. Yet few commendations are voiced for American public schools because US reading scores are highly competitive. Instead, public schools are castigated because mathematics scores are generally below average internationally.

A more balanced perspective is that overall (i.e., aggregating across four subjects as shown at the bottom of Figure 1), US students score somewhat higher than their peers in other industrialized nations, with only 24% of national scores being significantly higher than the US and 35% being significantly lower. Therefore, US students generally perform above average in international comparisons instead of poorly. There are good reasons for improving the effectiveness of American public education, but *not* because students perform poorly in comparison with their peers from other industrialized nations.

US Achievement by Grade Level

A critic of American public education observed that "In international comparisons, US students start out doing well in elementary grades and then fade by the end of high school" (Hanushek, 2002, p. 17), at which point US students outscored only Lithuania, Cyprus, and South Africa. Hanushek based this observation on TIMSS 1995 results in which he interpreted high-school performance to represent the achievement of students at age 17. This is a misconception, however, because the TIMSS 1995 tests were administered to students during the "final year of

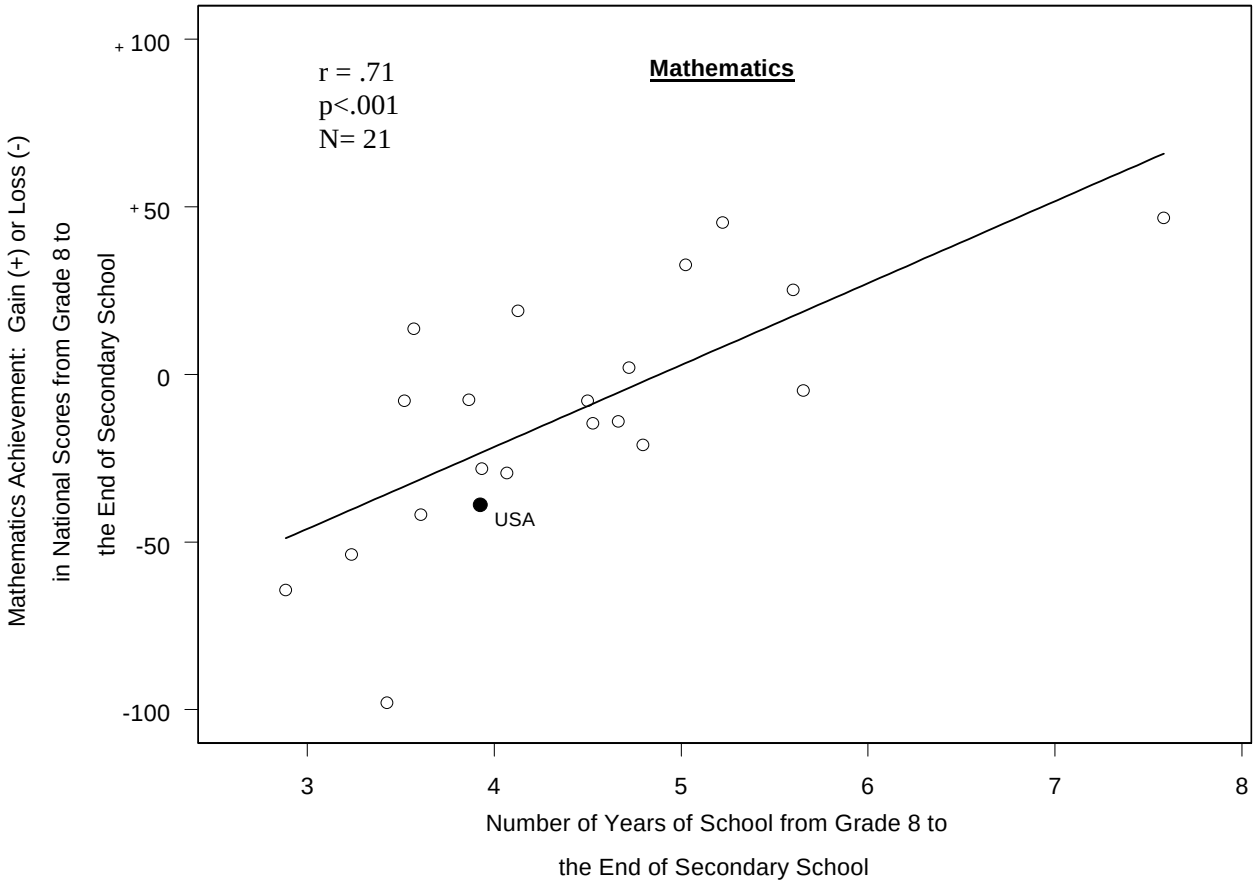

Figure 2. Changes in the relative standing of nations in mathematics achievement scores from grade 8 to the end of secondary school as a function of number of years of school between these grade levels. Each data point represents one nation. Improvements in achievement are shown as positive changes in national scores, while declines in achievement are shown as negative changes. Data source: TIMSS 1995.

secondary school" *as defined by each participating nation*. As shown in Figure 2, the number of years of schooling from grade 8 to the final year of secondary school varied widely across the 21 nations in the TIMSS sample (from about 3 to almost 8 years). When the gain or loss in national mean mathematics scores from grade 8 to the end of secondary school is plotted against the years of schooling past grade 8, we obtained a cross-national correlation of .71 (N = 21). Thus, the number of years of schooling beyond grade 8 is a major determinant of how well a nation performs at the end of secondary school. Since US students are disadvantaged in this comparison with only 4 years of schooling past grade 8, it is not appropriate to conclude from TIMSS 1995 data that the US performed poorly in international comparisons at the secondary level.

A better perspective on grade-level trends in US achievement in international comparisons is obtained by using all the international data included in Tables A-1 through A-7 (including the end of secondary school level from TIMSS 1995). The comparison of US achievement with other industrialized nations, aggregated across all subjects and surveys, is shown in Figure 3, separately at the elementary, middle, and secondary grades. As seen in Figure 3 and consistent with Hanushek's (2002) observation, US students started out doing very well at the elementary grades, and declined to average in the middle grades. It also appears that a further decline occurred from the middle to the secondary grades, with 45% of nations scoring significantly above the US, while 47% of the nations scored at an equivalent level to the US. The remaining 8% of nations scored significantly below the US. Viewed this way, the aggregated results make clear that the US did not perform "poorly" at the secondary level. Instead, the US performed somewhat below average at this level.

The percentages shown in Figure 3 for the secondary grades are skewed downward, however, by the inclusion of the "final year of secondary school" data from TIMSS 1995, as discussed above. If these data are eliminated, what remains is the PISA reading, mathematics, and science scores for grade 10 (see Table A-6). Peterson (2002) analyzed the PISA data and concluded that "the average combined score of US students in all three subjects falls at about the international average" (p. 46). In conclusion, it appears that US students score close to average in comparison with other industrialized nations at both the middle *and* secondary grades.

Nonetheless, the US decline in international competitiveness from the elementary to the middle and secondary grades is both a puzzle to understand and a problem for American public education to address. Whatever the cause, the observed declines in US student academic com-

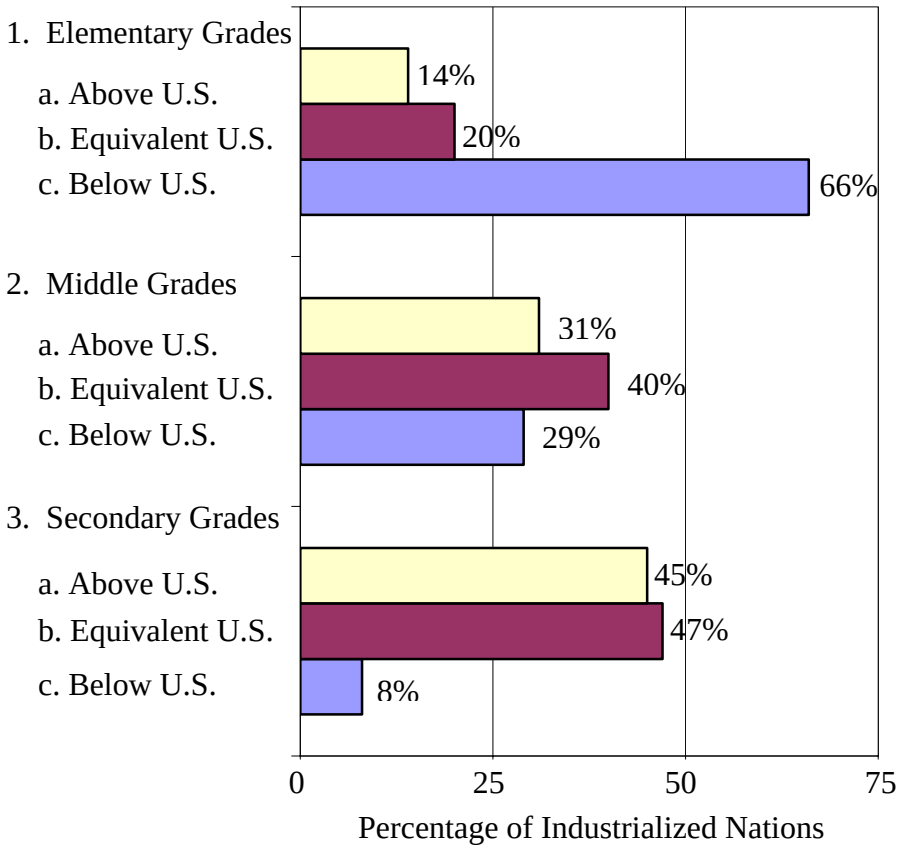

Figure 3. Percentage of 21 industrialized nations with national mean full-scale test scores significantly higher than, not significantly different than, and significantly lower than the comparable U.S. scores at the elementary (4^{th}), middle (8^{th} and 9^{th}), and secondary (10^{th} and end of secondary) grade levels across reading, mathematics, science, and civics. Sources: Five international education surveys sponsored by IEA and one by OECD.

petitiveness occurs over a period of years from grades 3 and 4 to grades 7 and 8 in TIMSS 1995. The decline cannot be explained by a number of important considerations because they are constant across grade levels from elementary to secondary. For example, the organization of the US education system is the same and financial resources are comparable, as is the socio-economic background and academic ability of students. Therefore, the explanation for the US decline must lie elsewhere. One of two major possibilities is that the quality of instruction declines in US middle and secondary schools relative to other nations. Another possibility is that student academic motivation declines in the US relative to other nations, perhaps as a byproduct of an adolescent peer culture that deflects attention from academic learning.

US Achievement Compared with G7 Nations

We conclude from the evidence reviewed above that US students do not perform *poorly* in academic achievement compared with other industrialized nations. Instead, they perform better than average at the elementary grades and average at the middle and secondary grades across six international surveys and four subjects. However, it is important to consider also how the US compares with the major industrialized nations of the world as defined by the Group of Seven (i.e., the G7) (Canada, France, Germany, Italy, Japan, UK, and US). After all, the list of 22 industrialized nations used for comparisons above includes a number of minor economic powers such as Iceland, New Zealand, Portugal, and Finland. The G7 nations are the major economic competitors of the US—the very nations that will cause economic decline if the performance of US students in mathematics and science is critical to the nation's future economic welfare (Colwell, 2000).

For achievement comparisons of the US with other G7 nations, we first compared US achievement scores by subject with those from each of the other six nations. For this purpose, comparisons were limited to surveys completed at the middle and secondary grades (i.e., grades 8, 9, and 10)—the grades at which the US was least competitive as shown in Figure 3. *Each* of the six other G7 nations was compared separately with the US on the surveys in which both nations participated. The results of these comparisons are shown by nation in Table A-8. These results were then aggregated for the five Western G7 nations (other than the US) and shown in Figure 4 in comparison with mean achievement scores for the US and Japan (the only nation from East Asia in the G7 group).

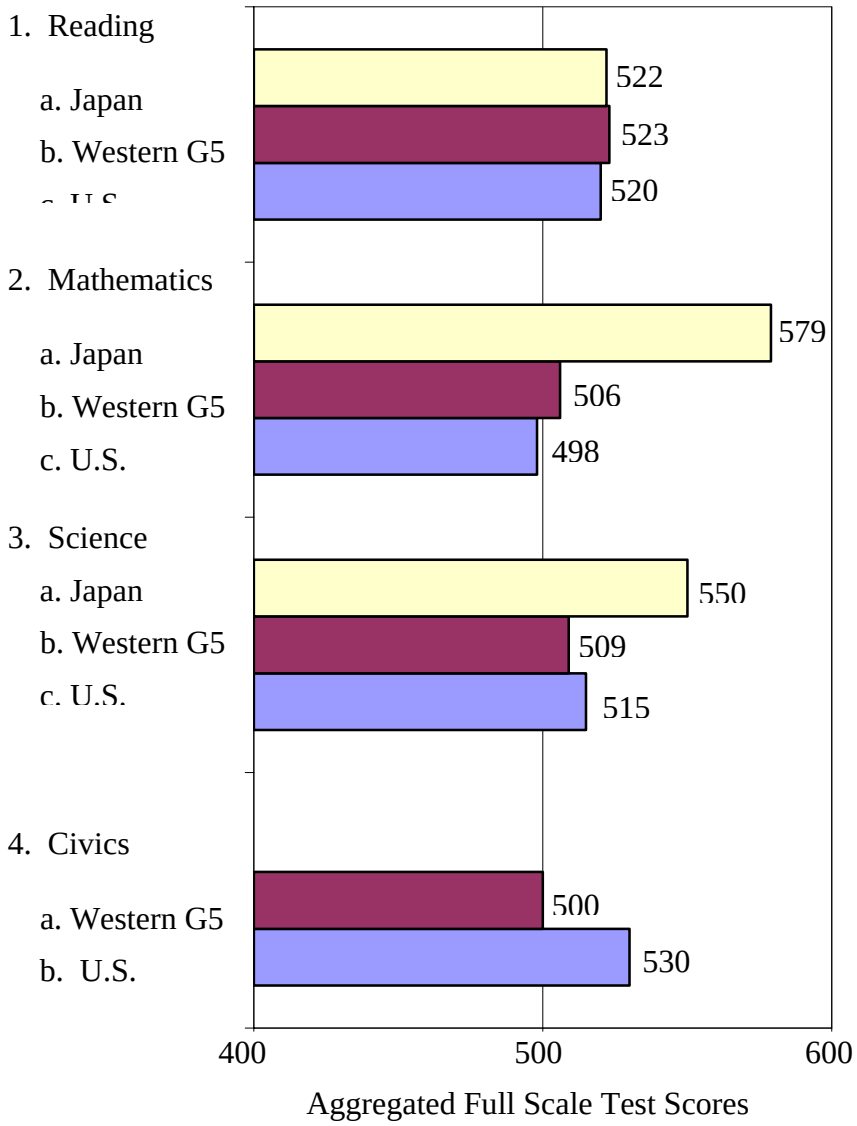

Figure 4. Aggregated national full-scale test scores of the U.S. in comparison with such scores for Japan and the median scores of the other Western G7 nations (Canada, England, France, Germany, and Italy) by subject matter at the 8th, 9th, and 10th grade levels combined. Civics scores were converted to a base of 500 for this figure. Sources: Five international education surveys sponsored by IEA and one by OECD.

As seen in Figure 4, US achievement is comparable to that of other Western G7 nations in reading, mathematics, and science, and significantly higher in civics (except for a tie with Italy; see also Table A-7). Scores of Japanese students in reading are comparable to Western G7 nations and the US, but much higher in mathematics and science (Japan did not participate in the Civics Education Study). With respect to academic achievement, the US is quite comparable to other major Western nations and should have little to fear in losing out economically (assuming a connection), whereas the Western nations consistently trail Japan in mathematics and science.[5]

Obviously there is a substantial difference in how well students from G7 Western nations perform in mathematics and science when compared with Japan as a representative of several East Asian nations (other consistently high performing nations from East Asia are Singapore, Korea, Chinese Taipei, and Hong Kong). It is a phenomenon of the West versus the East, not just the US versus the rest. As with the decline in competitiveness of US achievement scores from the elementary grades to the middle grades, the West-East differences in mathematics and science achievement are puzzles to be solved and a potential source of concern for the entire Western world. At least the US performs on a par with Western G7 nations.

US Diversity in Achievement Compared with G7 Nations

Not only does the US have the largest Gross Domestic Product among the G7 nations, but also it has by far the largest population, the largest number of partially autonomous states, and the most diverse population racially and ethnically. Given these many differences, how meaningful is it to compare US achievement scores with scores from much smaller and more homogeneous nations (such as England or Japan)? With respect to the racial composition of national populations, the U.S. Central Intelligence Agency (2000) provided recent estimates. The US is 86% White, whereas Germany and Italy are 100% White and Japan is 100% Asian. Similar to the US, Canada is 87% White. But in the US the largest minority is Black (12%) whereas in Canada it is Asian (12%). In addition, the US has a large ethnic minority (Hispanics). With respect to public

[5] The performance of Japanese students in mathematics and science can be compared with that of students from other G7 nations in TIMSS 1995, TIMSS 1999, and PISA. At the 8th and 10th grade levels, Japanese students scored significantly higher that students from all other G7 nations except for students from England in science in two of these surveys.

school enrollment in the US as of the fall 1996, 64% of students were White, 17% Black, 14% Hispanic, and 5% other (Asian, American Indian, etc.) (Snyder, Hoffman, & Geddes, 1999).

As is well known, there has long been a substantial gap in achievement scores between US White and Black students, and between White and Hispanic students. Accordingly, we analyzed international survey data to determine the extent to which the performance of US minority students might have impacted on US achievement scores in comparison with those of other G7 nations.

Specifically, analyses of US achievement scores for three racial/ethnic groups (White, Hispanic, Black) in comparison with G7 nations were limited to surveys completed at the middle and secondary grades (i.e., grades 8, 9, and 10)—the grades at which the US was least competitive. For this purpose, the national mean achievement scores for each racial/ethnic group were converted to percentile rank (PR) scores based on the group of industrialized nations included in each of Tables A-1 and A-3 through A-7. These PR scores were then used in making the paired national comparisons shown in Tables A-9 and A-10. Finally, these results were aggregated for the five Western G7 nations (other than the US) and compared with achievement scores for the US and Japan (the only nation from East Asia in the G7 group).

As seen in Figure 5, the US achievement scores for the majority White students were consistently greater than those of the other five Western G7 nations, even though these nations were predominantly White (Central Intelligence Agency, 2000). By comparison, the scores for US Black and Hispanic students were very low and well below other scores. This is compelling evidence that the low scores of two minority groups were major factors in reducing the comparative standing of the US in international achievement surveys. That is, if these minority students performed at the level of US White students, the US would lead all G7 nations (including Japan) in reading and would lead Western G7 nations in mathematics and science, while still trailing Japan in mathematics.[6]

Much has been written about the achievement disparities in the US, and progress toward closing the gap is a major objective of the educational reform strategies of the No Child Left Behind Act of 2001. Two perspectives on this problem predominate. One is that minority students perform poorly because the quality of their schooling is deficient. It is said that there are two

[6] Williams and Jocelyn (2000) report that US White and Japanese science scores from TIMSS 1995 at grade 8 were not significantly different.

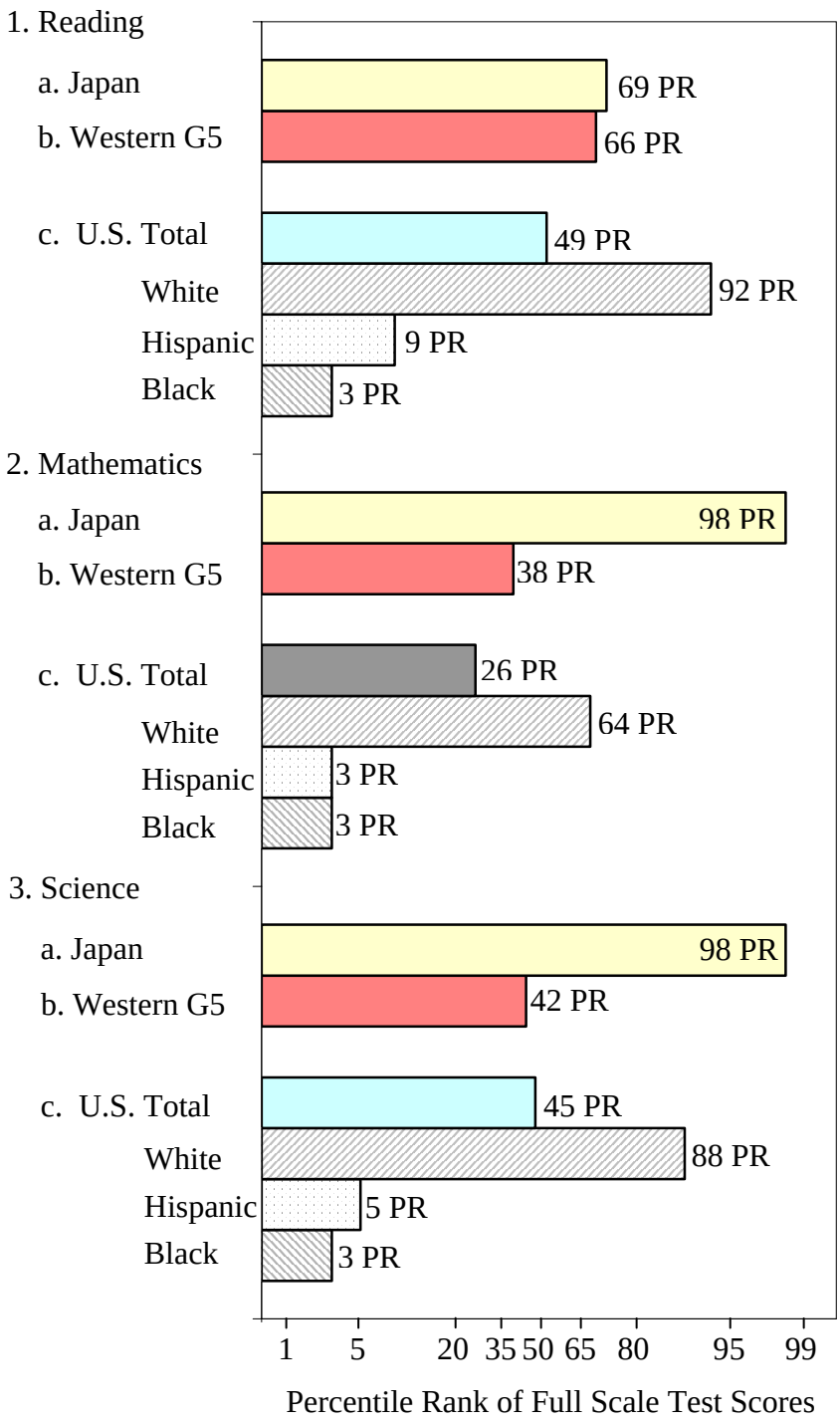

Figure 5. Percentile rank (PR) of aggregated national full-scale test scores for three ethnic/racial groups in the U.S. in comparison with such scores for Japan and median scores of other Western G7 nations (Canada, England, France, Germany, and Italy) by subject matter at the 8th, 9th, and 10th grade levels combined. Sources: Four international education surveys sponsored by IEA and one by OECD.

systems of American schools, a weak and poorly-funded system for our Black and Hispanic minorities and another that is academically strong and well-funded for the White majority (Levine, 2002; Bracey, 2002). The other perspective is that the achievement gap is driven by student background variables (i.e., economic disadvantage, limited parental support, discrimination) largely beyond the control of schools (Jencks, et al., 2002; Porter, 2003). No doubt there is some validity to both perspectives. If so, the problem continuing to face US public education is how to provide a type and quality of education that will compensate for the background disadvantages of minority students. If this can be achieved, the US may well realize its long-standing national goal of becoming first-in-the-world in mathematics and science achievement.

Conclusions

Based on our review of the results of six international achievement surveys on four subjects conducted over a recent 11-year period (1991-2001), we conclude that US students have generally performed *above average* in comparisons with industrialized nations instead of poorly. Certainly there is variability in performance, with the US scoring at a higher level in reading and civics than it does in mathematics and science. It also scored at a higher level in the elementary grades than in the middle and secondary grades. But even in mathematics and science at the middle and secondary levels, the US did not perform "poorly." Instead, the US scores were somewhat below average. If the US performed poorly, one would expect at least 75% of industrialized nations to have scored significantly higher. On the up side, US aggregated scores were above average in all subjects at the elementary levels, and in reading and civics across grade levels.

How then can the common perception be explained that the US generally performs poorly in the international achievement horse race? There are several plausible reasons for this, all of which might be partly correct.

First, it might be that many consumers of comparative education achievement statistics are simply not aware of the results of the full array of surveys in multiple subjects and grade levels that have been conducted in recent years. This may be due, in part, to the common emphasis on selective reporting of bad news by the press, and neglect of good news (Bracey, 1994). We hope the results of all international achievement comparisons reported here will provide more

comprehensive factual information on which anyone can base generally valid conclusions about the achievement of US students in international comparisons.

Second, it might be possible that others are aware of the achievement results of the full array of surveys recently completed, but simply view average results as "poor" because they expect the US to be first-in-the-world as it is in economic and military power. In this view, anything less than first place in the horse race is simply called a poor performance. Some might actually have taken seriously the political hype represented in the national goal adopted by state governors and President George H. W. Bush in 1990; viz., that "US students will be first-in-the-world in science and mathematics achievement" by the year 2000 (National Education Goals Panel, 1991, p. 16). Anything less would automatically qualify as poor performance.

This aspiration, however, is unreasonable. The US is not first-in-the-*industrialized*-world in minimizing the percentage of its population living in poverty or in minimizing infant mortality (Central Intelligence Agency, 2004). With such social and health indicators demonstrating that the US is not first-in-the-world, why should anyone expect the US to be first-in-the-world in educational achievement? There is, after all, abundant evidence that a number of social indicators are strongly associated with educational achievement.

Third, some observers might pick and choose from among existing surveys just the results for subjects and grades that support a preconception about the inadequacy of American public education, disregarding evidence to the contrary. The same can be said for both well-meaning reformers of public education and well-meaning conservatives who view public education as an untenable public monopoly that must be privatized as much and quickly as possible. The biased selection of evidence is not a surprising or uncommon strategy for advocates of a particular cause,[7] and may have strong policy and political impact. But to characterize the overall performance of US students as poor in international comparisons is not based on a balanced assessment of all the evidence; instead, it is misleading.

Is it reasonable to expect US student achievement in mathematics and science to improve in the next few decades so that the US scores substantially and consistently well above average? Regardless of current efforts to improve the outputs of public schools, it seems most unlikely.

[7] Weiss (1979) referred to this practice as the political model of research utilization.

At best, we can hope for incremental improvement such as that which appears to have occurred from international surveys before and after 1991.

As shown here, a major impediment to higher average achievement scores in the US is the performance of its US Black and Hispanic minorities. The achievement gap goes back decades, and is not closing rapidly (if at all). Moreover, demographic trends exacerbate the impact of the achievement gap on US mean achievement scores. In 1991, the population of public school students was 70% White and 28% Black and Hispanic (other minority was 4%). By the year 2000, Black and Hispanic students comprised about 34% of the student population, a stunning growth of 6% during a 10-year period (or more than 0.5% growth per year of these minority groups) (Snyder & Hoffman, 1993 & 2002). If the achievement gap remains constant, it is predictable that US mean scores will decline as the minority population increases as a percentage of the total.

The common way of looking at the international horse race in achievement may itself be misleading. That is, the practice of comparing nationally-representative samples of students regardless of national size, wealth, and racial and ethnic diversity. This practice leads to comparisons of the US with small nations such as Singapore and Finland, and with racially and ethnically homogeneous nations such as Iceland and Japan (Central Intelligence Agency, 2000). A more meaningful comparison would be between the United States of America and a hypothetical "United States of Europe" in which the mean scores for each state would be determined. These state-level scores from America and Europe could then be compared in their aggregate in terms of an overall average and the degree of variability.

As an illustration of this strategy, we prepared Figure 6 where student performance in science of US states is compared with the science performance of 22 European nations. The US state data come from 1996 NEAP for grade 8 (O'Sullivan, Reese, & Mazzeo, 1997), as linked to TIMSS by Johnson and Siegendorf (1998),[8] while the European data come from TIMSS 1995 for grade 8 (Peak, 1996). As seen in Figure 6, the distribution of US state-level mean science

[8] In 1996, 40 of 50 states participated voluntarily in the state NAEP examinations in science (Idaho, Illinois, Kansas, Nevada, New Hampshire, New Jersey, Ohio, Oklahoma, Rhode Island, and South Dakota did not participate)

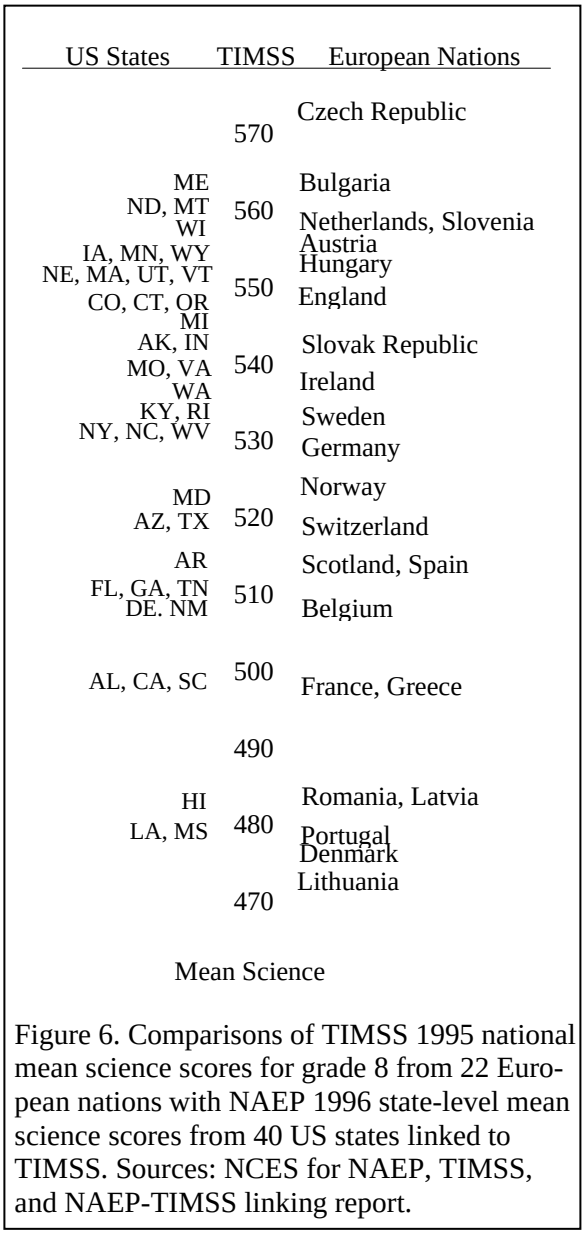

Figure 6. Comparisons of TIMSS 1995 national mean science scores for grade 8 from 22 European nations with NAEP 1996 state-level mean science scores from 40 US states linked to TIMSS. Sources: NCES for NAEP, TIMSS, and NAEP-TIMSS linking report.

scores is very similar to the distribution of mean science scores from 22 European nations.[9] At the extremes of the two distributions, the score for the Czech Republic (the highest scoring European nation) was not significantly higher than the 14 top scoring US states, while the scores for Louisiana and Mississippi (the two lowest scoring US states) were not significantly lower than those for the seven lowest scoring European nations. This simply illustrates that US states scored on a par with European nations in science in the mid-1990s (the only years for which comparable data are available for a most US states). Of all the nations from the East and the West that participated in the TIMSS 1995 science exams, only Singapore scored significantly higher than the 14 top scoring US states (Johnson & Siegendorf). Thus, the fairest of comparisons use US state-level achievement data. When this is done, the US performance in science is much like that of the rest of the industrialized world, including nations from the East.

In conclusion, this report should not be read as a defense of the status quo or an apology for inadequacies in US public education. There is always room for great improvement. In fact, all nations seem to be displeased with their education systems and levels of attainment. In the mid-1990s, the Organization for Economic Co-Operation and Development sponsored an international study of innovative methods in mathematics and science instruction in 23 nations (Black & Atkin, 1996). One of the interesting findings was that *none* of these nations (including some that performed well above average in the international horse race) was satisfied with mathematics and science teaching and learning in their nation. All were striving to improve, just as the US has been striving to improve—especially during the past two decades since the release of the federal "A Nation at Risk" report in 1983.

In recent years, the US has not performed "poorly" in international comparisons in a literal statistical sense. Consequently, an allegation of poor performance should not be used to tarnish the constructive work of the majority of public educators and the genuine attainments of achieving students. Nonetheless, the public and policymakers should continue to expect and demand improvements in instruction from educators and the educational system, and improvements in learning from students—especially in those subjects and grade levels where student achievement

[9] The mean score for each state and nation in Figure 6 is based on a sample of students. Accordingly, each score is subject to sampling error, and small to moderate differences among states and among nations are not statistically significant and should not be interpreted as meaningful. The significance of differences among science scores for European nations is reported by Beaton, Martin, et al. (1996), while the significance of differences between state science scores and scores for European nations is reported by Johnson and Siegendorf (1998). Figure 6 demonstrates that the distributions of science scores across states and across European nations are quite similar.

manifestly lags. Likewise, the public and policy makers should give credit, where due, and recognize genuine achievements in providing effective instruction by educators and in learning by students.

References

Baldi, S., Perie, M., Skidmore, D., Greenberg, E., & Hahn, C. (2001). *What democracy means to ninth-graders: U.S. results from the international IEA Civic Education Study.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Beaton, A.E., Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Smith, T.A., & Kelly, D.L. (1996). *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Boston: Boston College, Center for the Study of Testing, Evaluation and Educational Policy.

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Boston: Boston College, Center for the Study of Testing, Evaluation and Educational Policy.

Black, P., & Atkin, M. (1996). *Changing the subject: Innovations in science, mathematics and technology education*. London: Routledge.

Bracey, G.W. (1994). The media's myth of school failure. *Educational Leadership,52*(1), 80-83.

Bracey, G.W. (2002, January 23). International comparisons: an excuse to avoid meaningful educational reform [Commentary]. *Education Week*, pp 30, 32.

Central Intelligence Agency (2000). *The World Factbook, 2000.* Washington, DC: Brassey's Inc.

Central Intelligence Agency (2004). *The World Factbook, 2004.* Washington, DC: Brassey's Inc.

Colwell, R.R. (2000). National Science Foundation Director's Statement. In P. Gonzales, C. Calsyn, L. Jocelyn, K. Mak, D. Kastberg, S. Arafeh, T. Williams, & W. Tsen. (2000). *Pursuing Excellence: Comparisons of International Eighth-Grade Mathematics and Science Achievement from a U.S. Perspective, 1995 and 199*(vii-viii). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, K., Kastberg, D., Arafeh, S., Williams, T., & Tsen, W. (2000). *Pursuing excellence: Comparisons of international eighth-grade mathematics and science achievement from a U.S. perspective, 1995 and 1999—Initial findings form the Third International Math and Science Study-Repeat.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hanushek, E.A. (2002). The seeds of growth. *Education Next, Fall 2002*, 10-17.

Institute of Education Sciences. (2003). *Elective Mathematics Education Research Grants* (Request for Application Number NCER-03-05 of January 15, 2003). Washington, DC: Author.

International Monetary Fund Statistics Department. (2002). *International financial statistics* (Vol. 54). Washington, DC: Author.

Jaeger, R.M. (1994). Evaluating policy inferences drawn from international comparisons of students' achievement test performances. *Studies in Educational Evaluation, 20*(1), 23-39.

Jencks, C., Smith, M., Acland, H., Bane, M., Chen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.

Johnson, E.G., & Siegendorf, A. (1998). *Linking the National Assessment of Educational Progress and the Third International Mathematics and Science Study: Eighth-Grade Results* (NCES 98-500). Washington DC: U.S. Department of Education, National Center for Education Statistics.

Lemke, M., Calsyn, C., Lippman, L., Jocelyn, L, Kastberg, D., Liu, Y.Y., Roey, S., Williams, T., Kruger, T., & Bairu, G. (2001). *Outcomes of learning: Results from the 2000 Program for International Student Assessment of 15-year olds in reading, mathematics and science literacy*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Levine, A. (2002, September 12). *My old neighborhood* [Annual Report, President's Report]. Public Education Network Weekly NewsBlast. Retrieved September 12, 2002, from: http://www.tc.edu/annual/2002/essay.htm.

Martin, M.O., Mullis, I.V.S., Beaton, A.E., Gonzalez, E.J., Smith, T.A., & Kelly, D.L. (1997). *Science achievement in the primary years: IEA's Third International Mathematics and Science Study.* Boston: Boston College, Center for the Study of Testing, Evaluation and Educational Policy.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1997). *Mathematics achievement in the primary years: IEA's Third International Mathematics and Science Study.* Boston: Boston College, Center for the Study of Testing, Evaluation and Educational Policy.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Report.* Boston: Boston College, Center for the Study of Testing, Evaluation and Educational Policy.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., & Smith, T.A. (2000). *TIMSS 99 international mathematics report: Findings from the IEA's repeat of the Third International Mathematics and Science Study at the eighth grade.* Boston: Boston College, International Study Center, Lynch School of Education.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Smith, T.A., Chrostowski, S.J., Garden, R.A., O'Connor, K.M. (2000). *TIMSS 99 international science report: Findings from the IEA's repeat of the Third International Mathematics and Science Study at the eighth grade.* Boston: Boston College, International Study Center, Lynch School of Education.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Kennedy, A.M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary school in 35 countries.* Boston: Boston College, International Study Center, Lynch School of Education.

National Center for Education Statistics (2000). *Mathematics and science in the eighth grade: Findings from the Third International Mathematics and Science Study* (NCES 2000-014). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

National Center for Education Statistics, National Assessment of Educational Progress (NAEP (2000). *The nations report card: Mathematics, 2000.* Washington, DC: National Center for Education Statistics, U.S. Department of Education.

National Education Goals Panel. (1991). *The national education goals report: Building a nation of learners*. Washington, DC: Author.

Ogle, L., Sen, A., Pahlke, E., Jocelyn, L., Kastberg, D., Roey, S., & Williams, T. (2003). *International comparisons in fourth-grade reading literacy: Findings from the Progress in International Reading Study (PIRLS) of 2001.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Organisation for Economic Co-operation and Development. (2001). *Knowledge and skills for life: First results from PISA 2000.* Paris: Author.

O'Sullivan, C.Y., Reese, C.M., & Mazzeo, J. (1997). *NAEP 1996 Science report card for the nation and the states: Findings from the National Assessment of Educational Progress* (NCES 97-499). Washington, DC: National Center for Education Statistics.

Peak, L. (1996). *Pursuing excellence: A study of U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in international context.* Washington, DC.: National Center for Education Statistics, U.S. Department of Education.

Peterson, P.E. (2003). Ticket to nowhere. *Education next, 3*(2), 39-46.

Porter, A.C. (2003, October). *Prospects for school reform and closing the achievement gap*. Paper presented at Educational Testing Service's Invitational Conference, "Measurement and Research Issues in a New Accountability Era", New York.

Silver, E.A. (1998). *Improving mathematics in middle school: lessons from TIMSS and related research*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education

Snyder. T. D., & Hoffman, C. M. (1993). *Digest of education statistics: 1993*. Washington DC: National Center for Education Statistics, U.S. Department of Education.

Snyder, T. D., Hoffman, C. M., & Geddes, C. M. (1999). *Digest of education statistics: 1998.* Washington DC: National Center for Education Statistics, U.S. Department of Education.

Snyder. T. D., & Hoffman, C. M. (2002). *Digest of education statistics: 2001*. Washington DC: National Center for Education Statistics, U.S. Department of Education.

Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen.* Amsterdam: International Association for the Evaluation of Educational Achievement.

U.S. Department of Education, National Center for Education Statistics. (1994). *Reading literacy in the United States: Technical report.* Washington, DC: Author.

U.S. Department of Education, National Center for Education Statistics. (1996). *Reading literacy in the United States: Findings from the IEA Reading Literacy Study.* Washington, DC: Author.

Weiss, C.H. (1979). The many meanings of research utilization. *Public Administration Review, 39*, 426-431.

Williams, T., & Jocelyn, L. (2000). The distribution of student achievement. In U.S. Department of Education, National Center for Education Statistics (Ed.), *Mathematics and science in the eighth grade: Findings form the Third international Mathematics and Science Study* (pp. 49-84). Washington, DC: Author.

World Bank Development Economics Vice Presidency, Development Data Group. (2003). *World development indicators.* Washington, DC: World Bank, International Bank for Reconstruction and Development.

APPENDIX A

TABLES OF NATIONAL MEAN ACHIEVEMENT SCORES

AND COMPARISONS OF US WITH G7 NATIONS

Table A-1. Mean Reading Scores in Rank Order by Grade Level for Industrialized Nations Participating in the Reading Literacy Study (RLS 1991)

Grade 4			Grade 9		
Participating Industrialized Nation	Mean[a] Reading Score	Standard Error	Participating Industrialized Nation	Mean[a] Reading Score	Standard Error
Finland	569	3.4	Finland	560	2.5
United States[b]	547	2.8	France[b]	549	4.3
Sweden	539	2.8	Sweden	546	2.5
France[b]	531	4.0	New Zealand	545	5.6
Italy[b]	529	4.3	Iceland	536	0.0
New Zealand	528	3.3	Switzerland	536	3.2
Norway	524	2.6	**United States**[b]	535	4.8
Iceland	518	0.0	Denmark	525	2.1
MEAN: IND. NATIONS	515		*MEAN: IND. NATIONS*	524	
Switzerland	511	2.7	Portugal	523	3.1
Ireland	509	3.6	Germany (West)[b]	522	4.4
Belgium (French)	507	3.2	Norway	516	2.3
Greece	504	3.7	Italy[b]	515	3.4
Spain	504	2.5	Netherlands	514	4.9
Germany (West)[b]	503	3.0	Ireland	511	5.2
Netherlands	485	3.6	Greece	509	2.9
Portugal	478	3.6	Spain	490	2.5
Denmark	475	3.5	Belgium (French)	481	4.9

Source: USDE, NCES (1996)

[a]Mean scores for nations in dark shading above the U.S. were statistically significantly higher than the U.S. mean score; mean scores for nations in light shading were not significantly different than the U.S. mean score; mean scores for nations without shading below the U.S. were significantly lower than the U.S. mean score. This shading for statistical significance does not apply to the MEAN: INDUSTRIALIZED NATIONS.

[b]G-7 Nations.

Table A-2. Mean Reading Scores in Rank Order for Industrialized Nations Participating in the Progress in International Reading Literacy Study (PIRLS 2001)

	Grade 4	
Participating Industrialized Nation	Mean[a] Reading Score	Standard Error
Sweden	561	2.2
Netherlands	554	2.5
England[b]	553	3.4
United States[b]	542	3.8
Italy[b]	541	2.4
Germany[b]	539	1.9
MEAN: IND. NATIONS	534	
New Zealand	529	3.6
France[b]	525	2.4
Greece	524	3.5
Iceland	512	1.2
Norway	499	2.9

Source: Mullis, et al. (2003).

[a]Mean scores for nations in dark shading above the U.S. were statistically significantly higher than the U.S. mean score; mean scores for nations in light shading were not significantly different than the U.S. mean score; mean scores for nations without shading below the U.S. were significantly lower than the U.S. mean score. This shading for statistical significance does not apply to the MEAN: INDUSTRIALIZED NATIONS.

[b]G-7 Nations.

Table A-3. Mean Mathematics Scores in Rank Order by Grade Level for Industrialized Nations Participating in the Third International Mathematics and Science Study (TIMSS 2003)

Fourth Grade			Eighth Grade		
Participating Industrialized Nation	Mean[a] Math Score	Standard Error	Participating Industrialized Nation	Mean[a] Math Score	Standard Error
Japan	565	1.6	Japan	570	2.1
Netherlands	540	2.1	Netherlands	536	5.3
England	531	3.7	MEAN IND. NATIONS	507	
United States[b]	518	2.4	Australia	505	4.6
MEAN IND. NATIONS	514		**United States**[b]	504	3.3
Australia	499	3.9	Sweden	499	2.6
New Zealand	496	2.1	New Zealand	494	5.3
Norway	451	2.3	Italy	484	3.2
			Norway	461	2.5

Source: Gonzales et al. (2004).

[a]Mean scores for nations in dark shading above the U.S. were statistically significantly higher than the U.S. mean score; mean scores for nations in light shading were not significantly different than the U.S. mean score; mean scores for nations without shading below the U.S. were significantly lower than the U.S. mean score. This shading for statistical significance does not apply to the MEAN: INDUSTRIALIZED NATIONS.

[b] G-7 Nations.

Table A-4 New. Mean Science Scores in Rank Order by Grade Level for Industrialized Nations Participating in the Third International Mathematics and Science Study (TIMSS 2003)

Fourth Grade			Eighth Grade		
Participating Industrialized Nation	Mean[a] Math Score	Standard Error	Participating Industrialized Nation	Mean[a] Math Score	Standard Error
Japan	543	1.5	Japan	552	1.7
England	540	3.6	Netherlands	536	3.1
United States[b]	536	2.5	**United States**[b]	527	3.1
Netherlands	525	2.0	Australia	527	3.8
Australia	521	4.2	Sweden	524	2.7
MEAN IND. NATIONS	521		MEAN IND. NATIONS	521	
New Zealand	520	2.5	New Zealand	520	5.0
Italy	516	3.8	Norway	494	2.2
Norway	466	2.6	Italy	491	3.1

Source: Gonzales et al. (2004).

[a]Mean scores for nations in dark shading above the U.S. were statistically significantly higher than the U.S. mean score; mean scores for nations in light shading were not significantly different than the U.S. mean score; mean scores for nations without shading below the U.S. were significantly lower than the U.S. mean score. This shading for statistical significance does not apply to the MEAN: INDUSTRIALIZED NATIONS.

[b] G-7 Nations.

Table A-5. Mean Mathematics and Science Scores in Rank Order for Industrialized Nations Participating in the Third International Mathematics and Science Study (TIMSS 1999)

Mathematics Achievement: Grade 8			Science Achievement: Grade 8		
Participating Industrialized Nation	Mean[a] Score	Standard Error	Participating Industrialized Nation	Mean[a] Score	Standard Error
Japan[b]	579	1.7	Japan[b]	550	2.2
Netherlands	540	7.1	Netherlands	545	6.9
Canada[b]	531	2.5	Australia	540	4.4
Australia	525	4.8	England[b]	538	4.8
Finland	520	2.7	Finland	535	3.5
MEAN: IND. NATIONS	518		Canada[b]	533	2.1
United States[b]	502	4.0	*MEAN: IND. NATIONS*	529	
England[b]	496	4.1	**United States**[b]	515	4.6
New Zealand	491	5.2	New Zealand	510	4.9
Italy[b]	479	3.8	Italy[b]	493	3.9

Source: Mullis, Martin, Gonzalez, Gregory, Garden, et al. (2000) and Mullis, Martin, Gonzalez, Gregory, Smith, et al. (2000).

[a]Mean scores for nations in dark shading above the U.S. were statistically significantly higher than the U.S. mean score; mean scores for nations in light shading were not significantly different than the U.S. mean score; mean scores for nations without shading below the U.S. were significantly lower than the U.S. mean score. This shading for statistical significance does not apply to the MEAN: INDUSTRIALIZED NATIONS.

[b] G-7 Nations.

Table A-6. Mean Reading, Mathematics, and Science Scores in Rank Order for Industrialized Nations Participating in the Program for International Student Assessment (PISA 2003)

Reading Literacy: Grade 10			Mathematical Literacy: Grade 10			Scientific Literacy: Grade 10		
Participating Industrialized Nation	Mean[a] Score	Standard Error	Participating Industrialized Nation	Mean[a] Score	Standard Error	Participating Industrialized Nation	Mean[a] Score	Standard Error
Finland	543	1.6	Finland	544	1.9	Finland	548	1.9
Canada[b]	528	1.7	Netherlands	538	3.1	Japan[b]	548	4.1
Australia	525	2.1	Japan[b]	534	4.0	United Kingdom[b*]	532	2.7
United Kingdom[b*]	523	2.6	Canada[b]	532	1.8	Australia	525	2.1
New Zealand	522	2.5	Belgium	529	2.3	Netherlands	524	3.1
Ireland	515	2.6	Switzerland	527	3.4	New Zealand	521	2.4
Sweden	514	2.4	Australia	524	2.1	Canada[b]	519	2.0
Netherlands	513	2.9	New Zealand	523	2.3	Switzerland	513	3.7
Belgium	507	2.6	Iceland	515	1.4	France[b]	511	3.0
MEAN IND. NATIONS	502		Denmark	514	2.7	Belgium	509	2.5
Norway	500	2.8	France[b]	511	2.5	Sweden	506	2.7
Switzerland	499	3.3	Sweden	509	2.6	MEAN IND. NATIONS	506	
Japan[b]	498	3.9	United Kingdom[b*]	508	2.4	Ireland	505	2.7
France[b]	496	2.7	MEAN IND. NATIONS	507		Germany[b]	502	3.6
United States[b]	495	3.2	Austria	506	3.3	Iceland	495	1.5
Denmark	492	2.8	Germany[b]	503	3.3	**United States**[b]	491	3.1
Iceland	492	1.6	Ireland	503	2.4	Austria	491	3.4
Germany[b]	491	3.4	Norway	495	2.4	Spain	487	2.6
Austria	491	3.8	Spain	485	2.4	Italy[b]	486	3.1
Spain	481	2.6	**United States**[b]	483	2.9	Norway	484	2.9
Portugal	478	3.7	Portugal	466	3.4	Greece	481	3.8
Italy[b]	476	3.0	Italy[b]	466	3.1	Denmark	475	3.0
Greece	472	4.1	Greece	445	3.9	Portugal	468	3.5

Source: OECD (2004)

[a]Mean scores for nations in dark shading above the U.S. were statistically significantly higher than the U.S. mean score; mean scores for nations in light shading were not significantly different than the U.S. mean score; mean scores for nations without shading below the U.S. were significantly lower than the U.S. mean score. This shading for statistical significance does not apply to the MEAN: INDUSTRIALIZED NATIONS.

[b] G-7 Nations.

* Due to low response rates, the UK scores were not used in the NCES or OECD publications.

Table A-7. Mean Civics Scores in Rank Order for Industrialized Nations Participating in the Civic Education Study (CES 1999)

Participating Industrialized Nation[a]	Grade 9 Mean Civics Score	Grade 9 Standard Error
Finland	109	0.7
Greece	108	0.8
United States[b]	106	1.2
Italy[b]	105	0.8
Norway	103	0.5
Australia	102	0.8
MEAN: IND. NATIONS	102	
Denmark	100	0.5
Germany[b]	100	0.5
England[b]	99	0.6
Sweden	99	0.8
Switzerland	98	0.8
Portugal	96	0.7
Belgium (Fr.)	95	0.9

Source: Torney-Purta, et al. (2001)

[a]Mean scores for nations in dark shading above the U.S. were statistically significantly higher than the U.S. mean score; mean scores for nations in light shading were not significantly different than the U.S. mean score; mean scores for nations without shading below the U.S. were significantly lower than the U.S. mean score. This shading for statistical significance does not apply to the MEAN: INDUSTRIALIZED NATIONS.

[b] G-7 Nations.

Table A-8. U.S. Standing in Comparison with Other G-7 Nations in Mean Reading, Mathematics, Science and Civics Scores Aggregated Across the 8th, 9th, and 10th Grades.

	Aggregated National Mean Scaled Scores by Subject Matter			
Nation	Reading	Math	Science	Civics
England	523	506	538	99
vs. US	504	500	515	106
Germany	503	500	509	100
vs. US	520	497	517	106
France	527	528	499	-
vs. US	520	497	517	-
Italy	501	468	486	105
vs. US	520	498	507	106
Canada	534	531	531	-
vs. US	504	500	515	-
Japan	522	579	550	-
vs. US	504	500	515	-
G-6 Median	523	517	520	100
vs. US Median	512	499	515	106
Western G-5	523	506	509	100
vs. US Median	520	498	515	106

Note: The comparisons between the U.S. and other G-7 nations are based only on international survey data that were collected at the 8th, 9th, and 10th grade levels. Each paired comparison of the U.S. with another nation is based only on such survey data that were collected by both nations in the pair. The scaled scores listed here were aggregated by computing the median of all national mean scaled scores from the multiple grade levels and surveys that were thus classified in each cell of this table.

Table A-9. Comparison of U.S. with G-7 Nations England, Germany, and France in Mean Reading, Mathematics, Science and Civics Scores Aggregated Across the 8th, 9th, and 10 Grades by U.S. Ethnic/Racial Group

	Median Percentile Ranks (P.R.) of National Mean Scores[a]			
Nation	Reading	Math	Science	Civics
England	74	38	83	35
vs.				
US: Total[b]	36	26	45	81
US White	93	61	88	96
US Hispanic	2	3	6	19
US Black	2	3	3	4
Germany	28	35	37	42
vs.				
US: Total[b]	49	25	54	81
US White	92	64	91	96
US Hispanic	9	3	5	19
US Black	3	3	3	4
France	66	69	42	-
vs.				
US: Total[b]	49	25	54	-
US White	92	64	91	-
US Hispanic	9	3	5	-
US Black	3	3	3	-

Sources of Ethnicity/Race Achievement Data: for Reading, USDE, NCES (1994), Ogle, et al. (2003), Lemke et al. (2001); for Mathematics, USDE, NCES (2000), Gonzales et al. (2000), Lemke et al. (2001); for Science, USDE, NCES (2000), Gonzales et al. (2000), Lemke et al. (2001); for Civics, Baldi et al. (2001).

Note: The comparisons between the U.S. and other G-7 nations are based only on international survey data that were collected at the 8th, 9th, and 10th grade levels. Each paired comparison of the U.S. with another nation is based only on such survey data that were collected by both nations in the pair.

[a]The median percentile rank (P.R.) scores listed here are based on frequency distributions of national mean achievement scores of industrialized nations. They were computed as follows: national mean scaled scores were first converted to a P.R. score in the distributions of mean scores for industrialized nations shown in Tables A-1 and A-3 through A-7. Median P.R. scores were then computed from the several P.R. scores for multiple grade levels and surveys that were classified in each cell of this table. See the "Note" above. The P.R. scores for the mean scaled scores for U.S. ethnic/racial groups were similarly computed by substituting the mean scaled scores for each ethnic/racial group, in turn, for the U.S. Total mean scaled score. P.R. scores were rounded to the nearest whole number.

[b]The percentile rank for the U.S. Total was computed from the overall mean achievement scores that were based on all ethnic/racial groups (White, Hispanic, Black, Asian, and Other).

Table A-10. Comparison of U.S. with G-7 Nations Italy, Canada, and Japan in Mean Reading, Mathematics, Science and Civics Scores Aggregated Across the 8th, 9th and 10th Grades by U.S. Ethnic/Racial Group

	Median Percentile Ranks (P.R.) of National Mean Scores[a]			
Nation	Reading	Math	Science	Civics
Italy	25	9	9	73
vs.				
US: Total[b]	49	33	37	81
US White	92	68	86	96
US Hispanic	9	4	4	19
US Black	3	4	4	4
Canada	93	72	58	-
vs.				
US: Total[b]	36	26	45	-
US White	93	61	88	-
US Hispanic	2	3	6	-
US Black	2	3	3	-
Japan	69	98	98	-
vs.				
US: Total[b]	36	26	45	-
US White	93	61	88	-
US Hispanic	2	3	6	-
US Black	2	3	3	-

Sources of Ethnicity/Race Achievement Data: for Reading, USDE, NCES (1994), Ogle, et al. (2003), Lemke et al. (2001); for Mathematics, USDE, NCES (2000), Gonzales et al. (2000), Lemke et al. (2001); for Science, USDE, NCES (2000), Gonzales et al. (2000), Lemke et al. (2001); for Civics, Baldi et al. (2001).

Note: The comparisons between the U.S. and other G-7 nations are based only on international survey data that were collected at the 8th, 9th, and 10th grade levels. Each paired comparison of the U.S. with another nation is based only on such survey data that were collected by both nations in the pair.

[a]The median percentile rank (P.R.) scores listed here are based on frequency distributions of national mean achievement scores of industrialized nations. They were computed as follows: national mean scaled scores were first converted to a P.R. score in the distributions of mean scores for industrialized nations shown in Tables A-1 and A-3 through A-7. Median P.R. scores were then computed from the several P.R. scores for multiple grade levels and surveys that were classified in each cell of this table. See the "Note" above. The P.R. scores for the mean scaled scores for U.S. ethnic/racial groups were similarly computed by substituting the mean scaled scores for each ethnic/racial group, in turn, for the U.S. Total mean scaled score. P.R. scores were rounded to the nearest whole number.

[b]The percentile rank for the U.S. Total was computed from the overall mean achievement scores that were based on all ethnic/racial groups (White, Hispanic, Black, Asian, and Other).

APPENDIX B

DATA ANALYSIS METHODS

National mean full scale achievement data from Tables A-1 though A-7 were aggregated to make comparisons with US achievement data as shown in Figures 1 and 3 through 4. The aggregation methods are described below for each Figure.

Figure 1 Aggregation

Figure 1 includes comparisons of US achievement scores, *by subject matter*, with those of the 21 industrialized nations listed in Table 1. Achievement data from Tables A-1 through A-7 were aggregated *across* grade levels and surveys by the same method *separately* for each of four subjects (reading, mathematics, science, and civics). For each subject, the number of nations were counted that (a) scored significantly above the US, (b) did not score significantly higher or lower than the US, and (c) scored significantly below the US. Each of the three counts was then converted to a percentage of the their total, and reported in Figure 1.

Consider reading, for example. From Table A-1 it is seen that the US can be compared with 16 other nations at two grade levels for a total of 32 comparisons. From Table A-2, it is seen that the US can be compared with 10 other nations at one grade level. Finally, from Table A-6, it is seen that the US can be compared with 20 other nations in reading. The total number of comparisons across the three tables was therefore 62. Of these, 8 nations (or 13%) scored significantly above the US, 27 nations (or 44%) scored at a comparable level, and 27 nations (or 44%) scored significantly below the US. Thus aggregated, these percentages are shown in Figure 1 for reading.

The same procedure was used for mean national scores in mathematics, science, and civics to produce the percentages shown in Figure 1.

Figure 3 Aggregation

Figure 3 includes comparisons of US achievement scores, *by grade level*, with those of 21 industrialized nations. Achievement data from Tables A-1 through A-7 were aggregated *across* the four subjects and six surveys by the same method *separately* for each of three grade levels (elementary, middle, secondary). The elementary grades were represented by students age 9

years for reading, mathematics and science. The middle grades were represented by students age 12, 13, and 14 for reading, mathematics, science, and civics. The secondary grades were represented by students age 15 and 18 (or older) for reading, mathematics, and science. For each grade level, the number of nations were counted that (a) scored significantly above the US, (b) did not score significantly higher or lower than the US, and (c) scored significantly below the US. Each of the three counts was then converted to a percentage of their total, and reported in Figure 3.

Consider the elementary level, for example. From Table A-1 it is seen that the US can be compared with 16 other nations at the fourth grade. From Table A-2, it is seen that the US can be compared with 10 other nations at the fourth grade. Finally, from Tables A-4 and A-5, it is seen that the US can be compared with 12 other nations at the fourth grade for 24 comparisons. The total comparisons across the four tables was therefore 50. Of these, 7 nations (or 14%) scored significantly above the US, 10 nations (or 20%) scored at a comparable level, and 33 nations (or 66%) scored significantly below the US. Thus aggregated, these percentages are shown in Figure 3 for the elementary grades.

The same procedure was used for mean national scores at the middle and secondary grades to produce the percentages shown in Figure 3.

Figure 4 Aggregation

Figure 4 includes comparisons of US achievement scores, by subject matter, with those of the six nations of the G7 group listed in Table A-8. For this purpose, comparisons were limited to surveys completed at the middle and secondary grades (i.e., grades 8, 9, and 10)—the grades at which the US was least competitive as shown in Figure 3. TIMSS 1995 data for the "final year of secondary school" were not used because they did not represent a single grade level, but ranged from grade equivalents of 11 through 15 depending on the nation.

For Table A-8, achievement data from Tables A-1 through A-7 were aggregated *across* grade 8, 9, and 10 and across surveys, *separately* for each of the four subjects. *Each* of the six other G7 nations was first compared separately with the US on the surveys in which both paired nations participated (as shown in Table 1).

Consider England, for example. In reading, the US and England both participated in PISA but not in other reading surveys. The reading comparison was therefore limited to the PISA mean scores. It is these scores that are shown in the top section (for England vs. US) of Table A-8 in

the “Reading” column. In mathematics, the US and England participated in TIMSS 1995, TIMSS 1999, and PISA. For this comparison, the median of these three mean mathematics scores was computed for the US and compared with the median of these three mean mathematics scores for England. It is these median scores that are shown in the top section of Table A-8 in the “Mathematics” column. The same procedures were used to compute the science and civics comparison scores for the US and England as shown in the top section of Table A-8.

The procedures described above for the US versus England comparisons were also used for aggregating data for the comparisons of the US with each of the other G7 nations. The results are reported in Table A-8. By controlling this way for the national surveys in which each G7 nation participated, all paired comparisons of the US with each of the six other G7 nations were based on exactly the same survey results.

Next, the results for each individual comparison of the US with one other G7 nation (as shown in Table A-8) were aggregated across the other G7 nations. As shown at the bottom of Table A-8, medians of the other six G7 national aggregated scores (including Japan) and the other five Western G7 scores (excluding Japan) were computed, and then compared with the US median aggregated score for each country comparison shown in this table.

Finally, Figure 4 was based on the aggregated scores for Japan (from Table A-8), and the median scores for the Western G-5 nations (other than the US), and the US (both shown at the bottom of Table A-8). Because the Civics Survey scores were based on an overall mean of about 100, while the scores for the other three subjects were each based on an overall mean of about 500, the Civics Survey scores were multiplied by 5 to convert them to the same base as the other scores for the purpose of constructing Figure 4.

Figure 5 Aggregation

Figure 5 includes comparisons of US achievement scores for three racial/ethnic groups in the US (White, Hispanic, Black), by subject matter, with those of the six G7 nations listed in Tables A-9 and A-10. Comparisons were limited to surveys completed at the middle and secondary grades (i.e., grades 8, 9, and 10)—the grades at which the US was least competitive as shown in Figure 3. TIMSS 1995 data for the “final year of secondary school” were not used because they did not represent a single grade level, but ranged from grade equivalents of 11 through 15 depending on the nation.

For this purpose, the national mean achievement scores for grades 8, 9, and 10 (as shown in Tables A-1 and A-3 through A-7) were converted to percentile rank (PR) scores based on the group of industrialized nations included in each of these tables. With respect to each of these tables, the mean score for each of the three racial/ethnic groups was substituted (one at a time) for the US national mean score, and the PR score for each racial/ethnic group was computed based on the group of industrialized nations included in each of these tables. These PR scores were then used in making the paired national comparisons shown in Tables A-9 and A-10. In other respects, the aggregation of scores for Tables A-9 and A-10 was performed by the same procedures described above for Figure 4.

For Table A-9 and A-10, PR achievement data from Tables A-1 and A-3 through A-7 were aggregated *across* grades 8, 9, and 10 and across surveys, *separately* for each of the four subjects. PR scores for *each* of the six other G7 nations were first compared separately with each of the three racial/ethnic groups from the US on the surveys in which both paired nations participated (as shown in Table 1).

Consider England, for example. In reading, the US and England both participated in PISA but not in other reading surveys. The reading comparison was therefore limited to the PISA PR scores. It is these PR scores that are shown in the top section (for England vs. US Total and three racial/ethnic groups) of Table A-9 in the "Reading" column. In mathematics, the US and England participated in TIMSS 1995, TIMSS 1999, and PISA. For this comparison, the median of these three PR mathematics scores was computed for the US total and three racial/ethnic groups, and compared with the median of these three PR mathematics scores for England. It is these median PR scores that are shown in the top section of Table A-9 in the "Mathematics" column. The same procedures were used to compute the science and civics comparison PR scores for the US and England as shown in the top section of Table A-9.

The procedures described above for the US versus England comparisons were also used for aggregating data for the comparisons of the US with each of the other G7 nations. The results are reported in Tables A-9 and Table A-10. By controlling this way for the national surveys in which each G7 nation participated, all paired comparisons of the US total and three racial/ethnic groups, with each of the six other G7 nations were based on exactly the same the same survey results.

Finally, Figure 5 was based on (a) the aggregated PR scores for Japan (from Table A-10), (b) the median of 5 PR scores for the Western G-5 nations (other than the US) shown in Tables A-9 and A-10, and (c) the median of 5 PR scores for US total and three racial/ethnic groups (also shown in Tables A-9 and A-10) from each comparison with a Western G-5 nation. Civics PR scores were omitted from Figure 5 because three of the six Western G7 nations did not participate in the Civics Education Study.

Figure 6 NAEP-TIMSS Linking

The TIMSS 1995 mean science scores for grade 8 students from the 22 European nations were taken from Peak (1996), while Johnson and Steigendorf (1998) linked the NAEP 1996 mean science scores for 40 states with TIMSS 1995 science scores. The linkage reported for *each* of the 40 states were three groups of TIMSS 1995 nations: (a) those nations that scored significantly higher than each state, (b) those nations whose scores were not significantly different than each state, and (c) those nations that scored significantly lower than each state. As a point estimate of the linked TIMSS 1995 score for each state, we focused on the set of nations whose scores were *not* significantly different than the score for each state. We then computed the mean of the highest and lowest scoring nation in this set. The scores for these two nations were assumed to approximate the upper and lower limits of the confidence interval used by Johnson and Steigendorf to identify the set of nations whose scores were not significantly different than each linked state score. These TIMSS score estimates were then used to position each state in the order seen in Figure 6.