Linking NAEP Achievement Levels to TIMSS

Gary W. Phillips
Chief Scientist
American Institutes for Research

American Institutes for Research®

Expressing International Educational Achievement in Terms of U.S. Performance Standards: Linking NAEP Achievement Levels to TIMSS[1]

Gary W. Phillips
Chief Scientist
American Institutes for Research®
April 24, 2007

Introduction

Educators, researchers, and policymakers have considerable interest in how the American educational system compares to those in other countries. One major index for comparison is student academic achievement. Unfortunately, a lack of common metrics, as well as different definitions of performance standards, makes it difficult to compare measures of student achievement. The difficulty is similar to trying to compare the U.S. poverty level to that of other countries in the world. To do this, we first need a common metric. For example, we need to convert currencies of different countries to a common currency, such as dollars. Then we need a common definition and standard of poverty. That means either using a U.S. definition and standard and applying them to the rest of the world or using a common world definition and standard and applying those to the United States. No matter what common metric, definition, and standard are used, some people will argue it should have been done differently or not at all. This paper takes the position that such comparisons are not perfect, always require more research, and should be done with caution. However, such cross-country comparisons result in the cross-fertilization of information and help inform debate. In general, comparisons are useful in providing information to policymakers and the general public to help them achieve broad understandings that they otherwise would not have.

This paper links the scale of the *National Assessment of Educational Progress* (NAEP) to the scale of the *Third International Mathematics and Science Study* (TIMSS).[2] The purpose of this linking is to project the NAEP achievement levels onto the TIMSS scale. More specifically, the grade 8 NAEP: 2000 achievement levels in mathematics and science are projected on to the grade 8 TIMSS: 1999 assessment in mathematics and science. The linking equation is also applied to the 2003 TIMSS in mathematics and science. The goal is to project the grade 8 mathematics and science achievement levels in NAEP onto the TIMSS scale and thereby estimate the percent of basic, proficient, and advanced students in each country that participated in the 1999 TIMSS and 2003 TIMSS studies. The three achievement levels used were *basic, proficient*, and *advanced*, for both mathematics and science, as defined in *The Nation's Report Card: Mathematics 2000* (Braswell et al. 2001), and *The Nation's Report Card: Science 2000* (O'Sullivan et al. 2003), respectively. The TIMSS results may be found in *TIMSS 1999: International Mathematics Report* (Mullis et al. 2000), *TIMSS 1999: International Science Report* (Martin et al. 2000), *TIMSS 2003: International Mathematics Report* (Mullis et al. 2005), and *TIMSS 2003: International Science Report* (Martin et al. 2004).

[1] Copies of this paper can be downloaded by searching www.air.org and questions can be addressed to the author at gwphillips@air.org. Proper citation is as follows: Phillips, Gary W., *Expressing International Educational Achievement in Terms of U.S. Performance Standards: Linking NAEP Achievement Levels to TIMSS*, American Institutes for Research: Washington, DC, 2007.
[2] The definition of the acronym TIMSS was subsequently changed to Trends in International Mathematics and Science Study.

Linking Approaches

Mislevy (1992) and Linn (1993) have described many of the conceptual and statistical issues associated with linking assessments. They have outlined four forms of statistical linking: equating, calibration, projection, and statistical moderation. These are listed in descending order as a measure of their strength in linking. A more in depth discussion of linking is contained in the technical appendix.

In *equating,* both tests are designed and developed to be equally reliable, and each measures the same content. Equating is used when the goal is to relate two alternate forms of the same test, such as alternate forms of the ACT or the SAT.

In *calibration*, two tests are assumed to measure the same content, but they are not equally reliable. For example, one test might be a long test whereas the other is short. The two versions of the test are not equated, but they are indirectly comparable because they have been calibrated to a common scale. This type of linking is done across grades and across years in NAEP, TIMSS, most state criterion-referenced tests, and most nationally standardized, norm-referenced tests.

In *projection,* a regression equation uses the correlation between the two tests to predict the scores on one test from those of another test. There is no assumption that the two tests measure the same content or that they are equally reliable.

In *statistical moderation*, the scores on the first test are adjusted to have the same distributional characteristics as the scores on the second test. Statistical moderation does not use the correlation between the two tests.

Linking is essentially a process that provides a concordance table that expresses scores on one test (e.g., TIMSS) in terms of the metric of another test (e.g., NAEP). This paper uses *statistical moderation* to link the NAEP achievement levels to TIMSS by extending the process used in the *2000 NAEP–1999 TIMSS Linking Report* (Johnson et al. 2005). This extension was an extremely easy process because that report did all the hard work. The main goal of the report (Johnson et al. 2005) was to use the link between NAEP and TIMSS to estimate how the students in the states of the United States would have performed if they had taken the TIMSS test, based on the fact they took the NAEP test. This same linking process also can be used to answer the question, “How would other countries perform if their TIMSS results could be expressed in terms of NAEP achievement levels?” In other words, we can use the findings in the 2005 report by Johnson and colleagues to project the NAEP achievement levels onto the TIMSS scale as a way to interpret how each country performed on the TIMSS assessment in terms of U.S. performance standards. This paper takes that approach.

Linking NAEP to International Assessments

Several major attempts have been made to link NAEP statistically to international assessments.

The first attempt involved linking the 1991 International Assessment of Educational Progress (IAEP) to the 1992 NAEP in mathematics (Pashley and Phillips, 1993). The IAEP was first conducted in February 1988 in five countries (Ireland, Korea, Spain, the United Kingdom, and the United States) and four provinces in Canada (LaPointe, Mead, and Phillips, 1989) using representative samples of 13-year old students assessed in mathematics and science. The IAEP was expanded and repeated again in 1991 (LaPointe, Meade, and Askew, 1992) in 20 countries in which representative samples of 9- and 13-year old students were assessed in mathematics and science. Pashley and Phillips (1993) conducted the IAEP-NAEP linking study in mathematics using *projection* methodology. In order to establish the link between the IAEP and NAEP, a nationally representative linking sample of 1,609 students was administered both the IAEP and NAEP in 1992. The linking study used samples of 8th-grade students who took NAEP versus 13-year-old students who took the IAEP (NAEP was based on grade whereas the IAEP was based on age). The direction of the link was to predict NAEP performance from IAEP results in other countries. The purpose of the study was to estimate how other countries stacked up against the NAEP achievement levels. The IAEP-NAEP linkage was done within the context of the policy environment at the time. The nation's governors, along with the President had held the National Education Summit and adopted six broad national goals. The fourth goal was that, by the year 2000, "U.S. students would be the first in the world in science and mathematics achievement." The IAEP-NAEP linking study was the first effort to address directly the need for a common metric and common standard in international comparisons (i.e., predict how other countries would do on NAEP based on their performance on IAEP). Once the predicted NAEP scores were obtained, then the NAEP achievement levels were used to report different countries' performance. The IAEP was not repeated; however, it had many design features (such as linking studies) that were incorporated into subsequent international assessments of TIMSS.

A second attempt to link NAEP to an international study was done by Beaton and Gonzales (1993). They used *statistical moderation* to link the 1991 IAEP to the 1990 NAEP scale in mathematics. The results of the Beaton and Gonzales (1993) study were similar to the Pashley and Phillips (1993) study only for countries with performance similar to the U.S. average.

The third study used *statistical moderation* to link the grade 4 and grade 8 1996 NAEP to 1995 TIMSS, grades 4 and 8, mathematics and science (Johnson and Siengondorf, 1998). Based on the validation analyses (in two states that took both NAEP and TIMSS), the NAEP-TIMSS link appeared to work at grade 8 but not at grade 4.[3]

The fourth study (Johnson et al. 2005) used *projection* methods (similar to Pashley and Phillips, 1993) for grade 8 mathematics and science to link NAEP to TIMSS. The TIMSS assessment in mathematics and science was conducted in 1999, and the NAEP assessment in math and science was conducted in 2000. In addition to projection methods, the study also used *statistical moderation* as a secondary method of linking. Based on a validation study in which 12 states took both NAEP and TIMSS, the general finding was that, for the U.S. national linking sample, the projection method did not work. However, the statistical moderation method (which used the national samples of both NAEP and TIMSS instead of the linking sample) did perform well in the validation study.

Although statistical moderation provided an acceptable link, this approach is considered the weakest linking method because it does not use the correlation between the two assessments. In this case, however, it is the only method available so far that appears to work for linking NAEP to TIMSS. The estimates provided by statistical moderation should be considered rough, ballpark estimates and should be used only for broad policy understandings.

Purpose of this Paper

The main purpose of the NAEP-TIMSS link by Johnson and colleagues (2005) was to predict TIMSS results for the states within the United States, based on their performance on NAEP. The current paper uses the data and the formulas provided by that study to extend this process and link NAEP achievement

[3] The link worked at grade 8 based on the validation sample. The predicted TIMSS results for Minnesota (the only state that administered the 8th grade TIMSS) were comparable to the actual TIMSS results. The link did not work at grade 4. The predicted TIMSS results for the two states that administered 4th-grade TIMSS (Colorado and Minnesota) were considerably higher than the actual TIMSS results. The study was not able to determine why this result occurred in the grade 4 link.

levels to TIMSS. This analysis provides estimates of how countries outside the United States that participated in the TIMSS would perform, using the NAEP achievement levels estimated on the TIMSS scale.

Several important caveats are associated with these analyses. First, the standard errors and the validation analyses are based on data collected only within the United States. In the United States, students took both NAEP and TIMSS; in all other countries, however, students only took TIMSS. Whether the linking parameters are stable in other countries is an empirical question that the study by Johnson and colleagues (2005) could not answer. In fact, no international linking study has been designed to answer this question. There is no guarantee that linking parameters estimated from one group (e.g., the United States) will be the same in other groups.

The second caveat is that the percentage at or above basic, proficient, and advanced levels in the tables below is based on the assumption of a "normal distribution" of performance within each country. In most cases, this assumption should be approximately true.

The third caveat is that this paper used the linking parameters obtained from the 2000 NAEP and 1999 TIMSS to estimate achievement levels in the subsequent 2003 TIMSS; that is, the linking parameters are assumed to be stable across years. More than likely, they are not stable across years; nevertheless, they should be sufficient for very rough approximations. A better approach would be using a linking study that explicitly used the 2003 TIMSS. Because no linking study was conducted during the administration of the 2003 TIMSS, the past 1999–2000 study is all that is available. In fact, no linking studies have been conducted after the 2000 NAEP and 1999 TIMSS assessments.

Finally, the achievement levels developed for the NAEP were based on the content of the NAEP. Although similarities between the 8th-grade NAEP and TIMSS (Nohara, 2001) are substantial, the NAEP achievement levels do not strictly apply to TIMSS. The problem is similar to the poverty-level analogy used above. Definitions and standards of poverty in the United States will not strictly apply to other countries in the world; however, the definitions and standards can be used to estimate approximately how the rest of the world relates to U.S. expectations of a decent standard of living.

All of these caveats reinforce what was said above about the limits of inference from these data. At best, these concordance tables should be used for rough approximations and should not be used for less granular inferences.

Methodology

In the study by Johnson and colleagues (2005), NAEP was linked to TIMSS by using statistical moderation. This means the estimated $\hat{TIMSS}$ scores are actually NAEP scores adjusted to have the same mean and standard deviation as TIMSS. That is what it means in *statistical moderation* to say "NAEP is linked to TIMSS." The estimated $\hat{TIMSS}$ score associated with a NAEP achievement level ($\hat{TIMSS}_{level}$) is

$$\hat{TIMSS}_{level} = \hat{A} + \hat{B}\left(NAEP_{level}\right). \tag{1.1}$$

In equation (1.1) $\hat{A}$ is an estimate of the intercept of a straight line, and $\hat{B}$ is an estimate of the slope defined by

$$\hat{A} = \hat{\mu}_{TIMSS} - \hat{B}\hat{\mu}_{NAEP}$$
$$\hat{B} = \frac{\hat{\sigma}_{TIMSS}}{\hat{\sigma}_{NAEP}}. \qquad (1.2)$$

In equation (1.2), $\hat{\mu}_{NAEP}$ and $\hat{\mu}_{TIMSS}$ are the national means of the U.S. NAEP and TIMSS results for public school students, respectively, while $\hat{\sigma}_{NAEP}$ and $\hat{\sigma}_{TIMSS}$ are the standard deviations of the tests. The means and standard deviations in equation (1.2) are reported in table 1. The resulting estimates of the linking parameters $\hat{A}$ and $\hat{B}$ are reported in table 2.

Table 1 Means and standard deviations for national samples of grade 8 U.S. public school students, 1999 TIMSS and 2000 NAEP

Subject	TIMSS		NAEP	
	Mean	SD	Mean	SD
Mathematics	498.2	88.4	274.4	37.4
Science	510.4	98.0	149.2	36.2

SOURCES: National data file from the *1999 IEA Trends in International Mathematics and Science Study* (TIMSS-99) and the *2000 National Assessment of Educational Progress* (NAEP).

Table 2 Estimating 1999 TIMSS scores from 2000 NAEP, using statistical moderation with U.S. national samples

Subject	A	B
Mathematics	–150.38	2.36
Science	106.49	2.71

The NAEP achievement levels projected on to the TIMSS scale are reported in table 3 for mathematics and table 4 for science. The details of the estimation procedure for the standard error of the projected achievement levels are presented in excruciating detail in the technical appendix.

Table 3 Grade 8 2000 NAEP mathematics achievement levels linked to grade 8 1999-TIMSS mathematics

	NAEP achievement level	TIMSS Estimated achievement level	Standard error of TIMSS achievement level
Basic	262	469	4.83
Proficient	299	556	5.13
Advanced	333	637	6.72

Table 4 Grade 8 2000 NAEP science achievement levels linked to grade 8 1999-TIMSS science

	NAEP achievement level	TIMSS Estimated achievement level	Standard error of TIMSS achievement level
Basic	143	494	5.44
Proficient	170	567	5.59
Advanced	208	670	6.63

Results

The data presented in the tables below have important implications for policy because they pertain to efforts to improve U.S. achievement in mathematics and science. They shed additional light on comparisons between the United States and other countries and provide a useful application of NAEP achievement levels.

An ongoing problem in the analysis of international data is finding and using a common metric for international comparisons, particularly a metric with which many U.S. educators are familiar. In addition to overall average performance, using scaled scores, the common metric of the NAEP achievement levels is an important and easily understood measure of quality. That is, while states and countries can be ranked on an overall achievement score, linked information about the percentage of students predicted to be at or above basic, proficient, and advanced levels in other countries informs the analysis by providing more substantive comparisons. It also allows each state within the United States to compare the percentage of the state's students at each achievement level on NAEP with the percentage at and above each estimated achievement level on TIMSS in other countries.

The analyses in this paper provide a useful application of NAEP achievement levels. By projecting them onto the TIMSS scale, the NAEP achievement levels provide benchmarks for international comparisons.

Shortened versions of the content definitions of the 8th grade NAEP achievement levels in ***mathematics*** are provided in the NAEP 2000 mathematics report (Braswell et al. 2001, 8 and 11). The first sentence of the definitions is referred to as the policy definition of the achievement level.

> ***Basic*** level denotes partial mastery of the knowledge and skills that are fundamental for proficient work at a given grade. Eighth-grade students performing at the *Basic* level should exhibit evidence of conceptual and procedural understanding in the five NAEP content strands (number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions). This level of performance signifies an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percents.
>
> ***Proficient*** level represents solid academic performance. Students reaching this level demonstrate competency over challenging subject matter. Eighth-grade students performing at the *Proficient* level should apply mathematical concepts and procedures consistently to complex problems in the five NAEP content strands (number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions).

> ***Advanced*** level signifies superior performance at a given grade. Eighth-grade students performing at the *Advanced* level should be able to reach beyond the recognition, identification, and application of mathematical rules in order to generalize and synthesize concepts and principles in the five NAEP content strands (number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions).

The combination of the policy definitions and shortened versions of the content definitions of the 8th grade NAEP achievement levels in ***science*** are provided in the NAEP 2000 science report (O'Sullivan et al. 2003, 9 and 12).

> ***Basic*** level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade. Students performing at the *Basic* level demonstrate some of the knowledge and reasoning required for understanding of the Earth, physical, and life sciences at a level appropriate to grade 8. For example, they can carry out investigations and obtain information from graphs, diagrams, and tables. In addition, they demonstrate some understanding of concepts relating to the solar system and relative motion. Students at this level also have a beginning understanding of cause-and-effect relationships.
>
> ***Proficient*** level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter. Students performing at the *Proficient* level demonstrate much of the knowledge and many of the reasoning abilities essential for understanding of the Earth, physical, and life sciences at a level appropriate to grade 8. For example, students can interpret graphic information, design simple investigations, and explain such scientific concepts as energy transfer. Students at this level also show an awareness of environmental issues, especially those addressing energy and pollution.
>
> ***Advanced*** level signifies superior performance. Students performing at the *Advanced* level demonstrate a solid understanding of the Earth, physical, and life sciences as well as the abilities required to apply their understanding in practical situations at a level appropriate to grade 8. For example, students can perform and critique the design of investigations, relate scientific concepts to each other, explain their reasoning, and discuss the impact of human activities on the environment.

Before presenting the results it is important to understand how to interpret the tables that follow.

First, this report is a United States-oriented analysis that projects U.S. performance standards on to the TIMSS scale, then, statistically compares other counties to the United States. Although this analysis might help other countries interpret international results, it should be most helpful to the United States.

Second, the countries have been rank-ordered by percent estimated to be proficient in the tables that provide statistical comparisons (tables 5, 7, 10, and 12). The background calculations for these tables are carried out to many decimal places but have been rounded to the nearest whole number for the report. For example, in table 12, the U.S. and the Netherlands each report 31 percent estimated to be proficient. The United States is rank-ordered higher than the Netherlands because the U.S. percent estimated to be proficient is actually 31.20 percent, whereas the Netherlands percent estimated to be proficient is 30.73 (both are rounded to 31%).

Third, the rank-ordering has nothing to do with statistically significant differences. The rank-ordering was done to visually facilitate understanding but should not be used to do statistical comparisons to the United States. The pluses (+) and minuses (–) in the tables do this. As an example, in table 12, England (with 38 percent estimated to be proficient) is ranked higher than the United States (with 31 percent estimated to be proficient). However, when you take into account the margin of error in the survey, the two countries are not significantly different.

Finally, the statistical comparisons indicated by the pluses (+) and minuses (–) in tables 5, 7, 10, and 12 are comparisons between the United States and other countries. They do not apply to comparisons among other countries. For example, in table 10, let's say you wanted to see if the percent estimated to be proficient in Singapore (73%) is significantly different from Japan (57%). The difference would be significant if it was greater than, or less than, $1.96\sqrt{4.6^2 + 5.1^2} = 12.08$ (see the technical appendix for a discussion of the 95% confidence interval). Since the difference equals 16%, we can conclude that the percent estimated to be proficient in Singapore is significantly higher than Japan. However, comparisons like this that do not involve the United States, are not provided in table 10. For comparisons between all countries, see the technical appendix (for example, table 28 has the comparison between Singapore and Japan mentioned above).

Table 5 reports the projection of NAEP achievement onto the 1999 TIMSS grade-8 mathematics scale. Using the percentage at or above proficient as a benchmark, we see that 11 countries performed significantly better than the United States. Among them, five counties had more than twice the percentage of proficient students as the United States. These were Singapore; Republic of Korea; Hong Kong, SAR; Japan; and Chinese Taipei. These same countries had more than five times the percentage of advanced students. On the other hand, 17 countries' students performed significantly less well than those in the United States. The least proficient countries (those with single-digit proficiency percentages) in mathematics were Turkey, Indonesia, Islamic Republic of Iran, Tunisia, Chile, Philippines, Morocco, and South Africa.

Table 5 Percent of students at or above basic, proficient, and advanced in grade 8 1999-TIMSS mathematics: Estimated by linking the grade 8 2000 NAEP mathematics achievement levels to the grade 8 1999-TIMSS mathematics scale

Nation	Percent at or above basic	Margin of error for basic	Percent at or above proficient	Margin of error for proficient	Percent at or above advanced	Margin of error for advanced
Singapore	96+	1.7	73+	4.2	34+	4.9
Korea, Rep. of	93+	1.0	65+	2.7	26+	3.0
Hong Kong, SAR	94+	1.6	64+	3.9	23+	3.7
Japan	92+	1.1	61+	2.7	24+	2.7
Chinese Taipei	87+	1.6	61+	2.7	31+	2.9
Belgium (Flemish)	88+	1.9	51+	3.4	15+	2.6
Netherlands	83+	4.0	41+	5.5	9	3.2
Hungary	77+	2.5	39+	3.1	11	2.0
Slovak Republic	81+	2.7	38+	3.7	9	2.0
Slovenia	77+	2.3	38+	2.9	10	1.7
Canada	80+	2.3	36+	3.1	7	1.6
Russian Federation	75+	3.5	36	4.0	10	2.5
Australia	76+	3.2	35	3.7	8	2.1
Czech Republic	74+	3.1	32	3.4	7	1.8
Malaysia	73+	3.1	32	3.4	7	1.8
Bulgaria	69	3.7	30	3.7	7	2.0
Finland	78+	2.8	29	3.3	4	1.1
United States	**65**	**3.0**	**27**	**2.8**	**6**	**1.5**
Latvia (LSS)	68	3.0	26	2.8	5	1.2
England	63	3.2	23	2.8	5	1.3
New Zealand	60	3.6	23	3.0	5	1.5
Italy	55–	3.1	19–	2.3	3	1.0
Romania	51–	3.7	18–	2.8	4	1.3
Israel	49–	2.9	17–	2.1	4	1.0
Lithuania	57	3.7	17–	2.7	2–	1.0
Cyprus	53–	2.6	16–	1.7	3–	0.6
Moldova	50–	3.2	15–	2.2	2–	0.8
Thailand	49–	3.8	15–	2.5	2–	1.0
Macedonia, Rep. of	41–	3.0	12–	1.8	2–	0.7
Jordan	35–	2.4	11–	1.4	2–	0.6
Turkey	32–	3.1	7–	1.5	1–	0.5
Indonesia	26–	2.6	6–	1.4	1–	0.5
Iran, Islamic Rep.	29–	2.7	5–	1.1	0-	0.3
Tunisia	37–	3.4	5–	1.1	0–	0.2
Chile	18–	2.5	3–	0.9	0–	0.2
Philippines	10–	2.1	1–	0.8	0–	0.2
Morocco	7–	1.1	1–	0.3	0–	0.1
South Africa	4–	1.2	0–	0.4	0–	0.1

The nations have been rank ordered based on percent estimated to be proficient. The margin of error in the percentages for country *j* includes sampling error σ_{SEj} and linking error σ_{LEj}. The overall error is $\sigma_{Ej} = \sqrt{\sigma_{SEj}^2 + \sigma_{LEj}^2}$. A plus (+) or minus (–) indicates that we are 95% confident that the nation's percentage at and above the projected achievement level is greater or lesser than that in the United States.

One way of judging a nation's overall performance is to see how well the average student in that nation is performing on the projected NAEP achievement levels. If a nation's typical student (i.e., the nation's mean) is at or above the *proficient* level, then we might consider the nation to represent *world class* educational achievement. Using this criterion, we see in table 6 that only six nations met that standard in mathematics in 1999. *Unfortunately, the United States was not one of them*. If we use *below basic* as a criterion for nations that are clearly below the U.S. grade-level expectations, then almost one-third of the nations that participated in the study are performing below what we would expect in the United States. The lowest is South Africa, which had no students in the assessment functioning at the *proficient* level of achievement.

Table 6 Achievement levels associated with the national average in grade 8 1999-TIMSS mathematics
(basic = 469, proficient = 556, advanced = 637)

Nation	Mean	Level of nation's mean
Singapore	604	Proficient
Korea, Rep. of	587	Proficient
Chinese Taipei	585	Proficient
Hong Kong, SAR	582	Proficient
Japan	579	Proficient
Belgium (Flemish)	558	Proficient
Netherlands	540	Basic
Slovak Republic	534	Basic
Hungary	532	Basic
Canada	531	Basic
Slovenia	530	Basic
Russian Federation	526	Basic
Australia	525	Basic
Czech Republic	520	Basic
Finland	520	Basic
Malaysia	519	Basic
Bulgaria	511	Basic
Latvia (LSS)	505	Basic
United States	**502**	**Basic**
England	496	Basic
New Zealand	491	Basic
Lithuania	482	Basic
Italy	479	Basic
Cyprus	476	Basic
Romania	472	Basic
Moldova	469	Basic
Thailand	467	Below Basic
Israel	466	Below Basic
Tunisia	448	Below Basic
Macedonia, Rep. of	447	Below Basic
Turkey	429	Below Basic
Jordan	428	Below Basic
Iran, Islamic Rep	422	Below Basic

Nation	Mean	Level of nation's mean
Indonesia	403	Below Basic
Chile	392	Below Basic
Philippines	345	Below Basic
Morocco	337	Below Basic
South Africa	275	Below Basic

Table 7 reports similar results for the 1999 TIMSS in science. Only two nations—Chinese Taipei and Singapore—had a significantly higher percentage of proficient students than the United States. In science, 16 countries had significantly lower percentages of proficient students than in the United States. Using the average student compared to the projected NAEP proficient level of science achievement as a criterion, only two nations had *world class* educational achievement in science (table 8)—Chinese Taipei and Singapore.

Table 7 Percent of students at or above basic, proficient, and advanced in grade 8 1999-TIMSS science: Estimated by linking the grade 8 2000 NAEP science achievement levels to the grade 8 1999- TIMSS science scale

Nation	Percent at or above basic	Margin of error for basic	Percent at or above proficient	Margin of error for proficient	Percent at or above advanced	Margin of error for advanced
Chinese Taipei	80+	3.9	51+	5.5	13	3.5
Singapore	78+	4.7	51+	6.1	15+	4.3
Hungary	76+	4.4	43	5.6	8	2.6
Korea, Rep. of	74+	4.3	42	5.2	8	2.4
Japan	77+	4.4	41	5.8	6	2.1
Netherlands	75+	5.9	39	7.0	5	2.7
Australia	70	4.8	38	5.4	7	2.3
England	69	4.8	38	5.2	7	2.4
Czech Republic	71	5.1	36	5.7	5	2.1
Slovenia	68	4.9	34	5.1	5	1.9
Russian Federation	65	5.4	34	5.4	7	2.5
Finland	70	5.2	34	5.6	4	1.8
Slovak Republic	70	5.1	34	5.5	4	1.7
Canada	69	4.9	33	5.2	4	1.5
Belgium (Flemish)	73	5.5	32	6.1	3	1.3
Bulgaria	60	5.2	30	4.9	5	2.0
Hong Kong, SAR	70	5.8	30	5.9	2	1.3
United States	**59**	**4.9**	**30**	**4.5**	**6**	**1.9**
New Zealand	57	5.2	27	4.5	4	1.7
Latvia (LSS)	55	6.2	21	4.7	2	1.0
Italy	50	5.4	20	3.9	2	1.0
Malaysia	49	5.8	18	4.1	2	0.9
Israel	40–	4.5	17–	3.2	3	1.1
Lithuania	47	5.7	17–	3.8	1–	0.8
Romania	41–	5.2	16–	3.6	2	1.1
Macedonia, Rep. of	36–	4.8	13–	3.0	1–	0.8
Jordan	34–	4.2	13–	2.6	2–	0.7

Nation	Percent at or above basic	Margin of error for basic	Percent at or above proficient	Margin of error for proficient	Percent at or above advanced	Margin of error for advanced
Moldova	36–	4.6	13–	2.8	1–	0.7
Thailand	44	6.3	12–	3.5	1–	0.5
Cyprus	34–	4.9	10–	2.5	1–	0.4
Iran, Islamic Rep	29–	4.8	8–	2.3	0–	0.4
Indonesia	24–	4.6	6–	2.0	0–	0.3
Chile	20–	3.8	5–	1.5	0–	0.2
Turkey	22–	4.6	5–	1.8	0–	0.2
Philippines	11–	2.5	3–	1.3	0–	0.4
Tunisia	17–	4.5	2–	1.1	0–	0.1
Morocco	5–	1.4	1–	0.5	0–	0.1
South Africa	3–	1.1	1–	0.5	0–	0.1

The nations have been rank ordered based on percent estimated to be proficient. The margin of error in the percentages for country *j* includes sampling error σ_{SEj} and linking error σ_{LEj} . The overall error is $\sigma_{Ej} = \sqrt{\sigma_{SEj}^2 + \sigma_{LEj}^2}$. A plus (+) or minus (–) indicates that we are 95% confident that the nation's percentage at and above the projected achievement level is greater or less than that in the United States.

Table 8 Achievement levels associated with the national average in grade 8 1999-TIMSS science (basic = 494, proficient = 567, advanced = 670)

Nation	Mean	Level of nation's mean
Chinese Taipei	569	Proficient
Singapore	568	Proficient
Hungary	552	Basic
Japan	550	Basic
Korea, Rep. of	549	Basic
Netherlands	545	Basic
Australia	540	Basic
Czech Republic	539	Basic
England	538	Basic
Finland	535	Basic
Slovak Republic	535	Basic
Belgium (Flemish)	535	Basic
Slovenia	533	Basic
Canada	533	Basic
Hong Kong, SAR	530	Basic
Russian Federation	529	Basic
Bulgaria	518	Basic
United States	**515**	**Basic**
New Zealand	510	Basic
Latvia (LSS)	503	Basic
Italy	493	Below Basic
Malaysia	492	Below Basic
Lithuania	488	Below Basic
Thailand	482	Below Basic

Nation	Mean	Level of nation's mean
Romania	472	Below Basic
Israel	468	Below Basic
Cyprus	460	Below Basic
Moldova	459	Below Basic
Macedonia, Rep. of	458	Below Basic
Jordan	450	Below Basic
Iran, Islamic Rep	448	Below Basic
Indonesia	435	Below Basic
Turkey	433	Below Basic
Tunisia	430	Below Basic
Chile	420	Below Basic
Philippines	345	Below Basic
Morocco	323	Below Basic
South Africa	243	Below Basic

When looked at through the lens of projected NAEP achievement levels, the general picture that emerges for science is that students in the participating countries do not do as well in science as they do in mathematics. However, this conclusion may be a non sequitur; the "bar" for the projected NAEP achievement levels in science is probably higher than in mathematics. Evidence for this conclusion can be found by comparing the TIMSS international benchmarks to the projected NAEP achievement levels. The four TIMSS international benchmarks developed in the 2003 TIMSS in grades 4 and 8 are: advanced (625), high (550), intermediate (475), and low (400). The international benchmarks are the same for both mathematics and science and are comparable from a normative point of view. Because the projected NAEP achievement levels are on the same scale as TIMSS, they can be compared to the international benchmarks. These comparisons are presented in table 9.

Table 9 TIMSS international benchmarks compared to projected NAEP achievement levels

TIMSS	TIMSS international benchmarks	NAEP	Projected NAEP achievement level in math	Projected NAEP achievement level in science	Projected NAEP achievement level minus TIMSS international benchmark in mathematics	Projected NAEP achievement level minus TIMSS international benchmark in science
Advanced	625	Advanced	637	670	12	45
High	550	Proficient	556	567	6	17
Intermediate	475	Basic	469	494	–6	19
Low	400					

The projected NAEP achievement levels in mathematics are actually close to the international benchmarks. However, all three of the projected NAEP achievement levels in science are higher than the international benchmarks. In fact, the projected advanced NAEP science achievement level is substantially higher and is almost one-half of a standard deviation above the international advanced benchmark (the international standard deviation in TIMSS is equal to 100).

In 2003, the TIMSS survey was expanded from 38 nations to 46 nations, bringing into the survey a few more mostly underachieving countries. In 2003, there were five countries with significantly more proficient mathematics students than the United States. Furthermore, the same five countries that were ranked highest achieving in mathematics in 1999 (with twice the percentage of proficient students) were the highest achieving again. In table 10, we see these were Singapore; Hong Kong, SAR; Republic of Korea; Chinese Taipei; and Japan. Even more significant was the percentage of advanced students in these five countries. Each of these countries had four to seven times the percentage of advanced students as the United States. There were 19 counties which were significantly below the United States in their percentages of proficient students. These were the Republic of Moldova, Cyprus, Norway, the Republic of Macedonia, Jordan, Egypt, Indonesia, Palestinian National Authority, Islamic Republic of Iran, Chile, Bahrain, Philippines, Tunisia, Morocco, Botswana, Saudi Arabia, Ghana, and South Africa. Four nations had no one in the TIMSS assessment functioning at the proficient level. These nations were Botswana, Ghana, Saudi Arabia, and South Africa.

Table 10 Percentage of students at or above basic, proficient, and advanced in grade 8 2003-TIMSS mathematics: Estimated by linking the grade 8 2000 NAEP mathematics achievement levels to the grade 8 1999-TIMSS mathematics scale

Nation	Percent at or above basic	Margin of error for basic	Percent at or above proficient	Margin of error for proficient	Percent at or above advanced	Margin of error for advanced
Singapore	96+	1.5	73+	4.6	35+	6.4
Hong Kong, SAR	95+	1.7	66+	5.5	24+	6.0
Korea, Rep. of	92+	1.8	65+	4.6	29+	5.4
Chinese Taipei	88+	2.4	61+	4.5	30+	5.0
Japan	90+	2.3	57+	5.1	20+	4.7
Belgium (Flemish)	82+	3.7	40	5.6	9	3.0
Netherlands	83+	4.0	38	6.2	7	3.0
Hungary	77	3.9	37	5.1	9	2.9
Estonia	82+	4.0	36	5.8	6	2.6
Slovak Republic	68	4.5	28	4.5	6	2.1
Australia	67	4.9	27	4.7	5	2.2
Russian Federation	69	4.8	27	4.8	5	2.0
Malaysia	70	5.1	26	5.0	4	1.9
United States	**67**	**4.7**	**26**	**4.4**	**5**	**1.9**
Latvia	70	4.9	25	4.8	4	1.8
Lithuania	66	4.7	24	4.3	4	1.7
Israel	63	4.6	24	4.0	5	1.8
England	65	5.4	22	4.7	4	1.8
Scotland	65	5.2	22	4.4	3	1.5
New Zealand	63	5.6	21	4.7	3	1.8
Sweden	66	5.2	21	4.3	3	1.3
Serbia	54–	4.5	19	3.2	4	1.3
Slovenia	63	5.2	19	4.0	2	1.1
Romania	53–	5.0	18	3.6	4	1.5
Armenia	54	4.8	18	3.4	3	1.2
Italy	58	5.2	17	3.7	2	1.2
Bulgaria	53	5.2	17	3.6	3	1.3
Moldova, Rep. of	46–	5.2	12–	2.9	1	0.9
Cyprus	45–	4.7	11–	2.5	1	0.6

Nation	Percent at or above basic	Margin of error for basic	Percent at or above proficient	Margin of error for proficient	Percent at or above advanced	Margin of error for advanced
Norway	46–	5.6	9–	2.5	1–	0.5
Macedonia, Rep. of	35–	4.4	8–	2.1	1	0.6
Jordan	31–	4.3	7–	1.9	1–	0.5
Egypt	25–	3.6	5–	1.4	1–	0.4
Indonesia	26–	4.2	5–	1.7	1–	0.5
Palestinian Nat'l. Auth.	20–	3.1	4–	1.1	0–	0.3
Lebanon	30–	5.3	3–	1.4	0–	0.2
Iran, Islamic Rep. of	22–	4.0	2–	0.9	0–	0.1
Chile	16–	3.2	2–	0.8	0–	0.2
Bahrain	19–	3.4	2–	0.7	0–	0.1
Philippines	15–	3.3	2–	1.0	0–	0.2
Tunisia	16–	4.1	1–	0.5	0–	0.0
Morocco	11–	2.9	1–	0.4	0–	0.0
Botswana	8–	2.1	0–	0.3	0–	0.0
Saudi Arabia	3–	1.0	0–	0.3	0–	0.1
Ghana	4–	1.6	0–	0.3	0–	0.0
South Africa	2–	0.8	0–	0.2	0–	0.0

The nations have been rank ordered based on percent estimated to be proficient. The margin of error in the percentages for country *j* includes sampling error σ_{SEj} and linking error σ_{LEj}. The overall error is $\sigma_{Ej} = \sqrt{\sigma_{SEj}^2 + \sigma_{LEj}^2}$. A plus (+) or minus (–) indicates that we are 95% confident that the nation's percentage at and above the projected achievement level is greater or less than that in the United States.

Table 11 Achievement levels associated with the national average in grade 8 2003-TIMSS mathematics
(basic = 469, proficient = 556, advanced = 637)

Nation	Mean	Level of nation's mean
Singapore	605	Proficient
Korea, Rep. of	589	Proficient
Hong Kong, SAR	586	Proficient
Chinese Taipei	585	Proficient
Japan	570	Proficient
Belgium (Flemish)	537	Basic
Netherlands	536	Basic
Estonia	531	Basic
Hungary	529	Basic
Slovak Republic	508	Basic
Russian Federation	508	Basic
Malaysia	508	Basic
Latvia	508	Basic
Australia	505	Basic
United States	**504**	**Basic**
Lithuania	502	Basic

Nation	Mean	Level of nation's mean
Sweden	499	Basic
England	498	Basic
Scotland	498	Basic
Israel	496	Basic
New Zealand	494	Basic
Slovenia	493	Basic
Italy	484	Basic
Armenia	478	Basic
Serbia	477	Basic
Bulgaria	476	Basic
Romania	475	Basic
Norway	461	Below Basic
Moldova, Rep. of	460	Below Basic
Cyprus	459	Below Basic
Macedonia, Rep. of	435	Below Basic
Lebanon	433	Below Basic
Jordan	424	Below Basic
Indonesia	411	Below Basic
Iran, Islamic Rep. of	411	Below Basic
Tunisia	410	Below Basic
Egypt	406	Below Basic
Bahrain	401	Below Basic
Palestinian Nat'l Auth.	390	Below Basic
Chile	387	Below Basic
Morocco	387	Below Basic
Philippines	378	Below Basic
Botswana	366	Below Basic
Saudi Arabia	332	Below Basic
Ghana	276	Below Basic
South Africa	264	Below Basic

Table 12 shows that two nations had a significantly higher percentage of students proficient in science than the United States. Twenty-five nations had a smaller percentage of proficient students than the United States. Two nations, Singapore and Chinese Taipei, had students whose average performance was at the proficient level in science (table 13).

Table 12 Percent of students at or above basic, proficient, and advanced in grade 8 2003-TIMSS science: Estimated by linking the grade 8 2000 NAEP science achievement levels to the grade 8 1999-TIMSS science scale

Nation	Percent at or above basic	Margin of error for basic	Percent at or above proficient	Margin of error for proficient	Percent at or above advanced	Margin of error for advanced
Singapore	82+	3.5	55+	5.2	16+	3.8
Chinese Taipei	84+	3.7	52+	5.9	11	3.3
Korea, Rep. of	82+	4.1	45	6.3	6	2.2
Hong Kong, SAR	83+	4.5	44	6.9	4	2.0
Japan	79+	4.4	42	6.1	5	1.9
Estonia	82+	4.6	41	6.8	4	1.7
England	74	5.0	38	6.0	5	2.1
Hungary	74	4.8	38	5.7	5	1.9
United States	**66**	**5.1**	**31**	**5.1**	**4**	**1.6**
Netherlands	76	5.9	31	6.7	1	1.0
Australia	67	5.6	30	5.6	3	1.4
Sweden	66	5.5	28	5.3	2	1.2
New Zealand	64	6.3	26	5.7	2	1.3
Slovak Republic	62	5.7	26	5.0	2	1.1
Lithuania	64	5.9	25	5.1	2	0.8
Slovenia	65	6.0	24	5.2	1	0.7
Russian Federation	61	6.0	24	5.0	2	1.1
Scotland	60	5.8	24	4.8	2	1.0
Belgium (Flemish)	63	6.2	22	5.1	1	0.7
Latvia	61	6.4	21	4.9	1	0.6
Malaysia	60	6.8	20	5.1	1	0.7
Israel	47–	5.3	18–	3.6	2	0.8
Bulgaria	44–	5.3	17–	3.7	2	1.0
Italy	49–	5.8	17–	3.8	1	0.6
Jordan	42–	5.1	15–	3.3	1	0.7
Norway	50	6.3	15–	3.8	1–	0.4
Romania	40–	5.2	14–	3.3	1	0.8
Serbia	38–	5.0	12–	2.8	1	0.4
Macedonia, Rep. of	31–	4.5	10–	2.4	1	0.5
Moldova, Rep. of	39–	5.9	10–	3.0	0–	0.3
Armenia	34–	5.2	10–	2.6	1–	0.4
Egypt	24–	3.6	8–	1.9	1	0.4
Palestinian Nat'l. Auth.	26–	4.1	8–	1.9	1–	0.3
Iran, Islamic Rep. of	29–	5.2	6–	1.9	0–	0.2
Cyprus	25–	4.4	6–	1.7	0–	0.2
Bahrain	23–	4.4	4–	1.4	0–	0.1
Chile	17–	3.4	3–	1.2	0–	0.1
Indonesia	18–	4.0	3–	1.3	0–	0.2
Philippines	13–	2.9	3–	1.3	0–	0.3
Lebanon	14–	3.0	3–	1.2	0–	0.2
Saudi Arabia	9–	2.9	1–	0.7	0–	0.1
Botswana	7–	1.8	1–	0.5	0–	0.0
South Africa	3–	1.0	1–	0.5	0–	0.1

Nation	Percent at or above basic	Margin of error for basic	Percent at or above proficient	Margin of error for proficient	Percent at or above advanced	Margin of error for advanced
Morocco	8–	2.5	1–	0.4	0–	0.0
Ghana	2–	0.9	0–	0.4	0–	0.1
Tunisia	7–	2.5	0–	0.3	0–	0.0

The nations have been rank ordered based on percent estimated to be proficient. The margin of error in the percentages for country *j* includes sampling error σ_{SEj} and linking error σ_{LEj}. The overall error is $\sigma_{Ej} = \sqrt{\sigma_{SEj}^2 + \sigma_{LEj}^2}$. A plus (+) or minus (–) indicates that we are 95% confident that the nation's percentage at and above the projected achievement level is greater or less than that in the United States.

Table 13 Achievement levels associated with the national average in grade 8 2003-TIMSS science (basic = 494, proficient = 567, advanced = 670)

Nation	Mean	Level of nation's mean
Singapore	578	Proficient
Chinese Taipei	571	Proficient
Korea, Rep. of	558	Basic
Hong Kong, SAR	556	Basic
Japan	552	Basic
Estonia	552	Basic
England	544	Basic
Hungary	543	Basic
Netherlands	536	Basic
United States	**527**	**Basic**
Australia	527	Basic
Sweden	524	Basic
New Zealand	520	Basic
Slovenia	520	Basic
Lithuania	519	Basic
Slovak Republic	517	Basic
Belgium (Flemish)	516	Basic
Russian Federation	514	Basic
Scotland	512	Basic
Latvia	512	Basic
Malaysia	510	Basic
Norway	494	Basic
Italy	491	Below Basic
Israel	488	Below Basic
Bulgaria	479	Below Basic
Jordan	475	Below Basic
Moldova, Rep. of	472	Below Basic
Romania	470	Below Basic
Serbia	468	Below Basic
Armenia	461	Below Basic
Iran, Islamic Rep. of	453	Below Basic
Macedonia, Rep. of	449	Below Basic

Nation	Mean	Level of nation's mean
Cyprus	441	Below Basic
Bahrain	438	Below Basic
Palestinian Nat'l Auth.	435	Below Basic
Egypt	421	Below Basic
Indonesia	420	Below Basic
Chile	413	Below Basic
Tunisia	404	Below Basic
Saudi Arabia	398	Below Basic
Morocco	396	Below Basic
Lebanon	393	Below Basic
Philippines	377	Below Basic
Botswana	365	Below Basic
Ghana	255	Below Basic
South Africa	244	Below Basic

Summary and Recommendations

Education policymakers struggle every day with trying to make sense out of national and international data. One big problem that makes understanding difficult for a U.S. audience is that assessments conducted internationally (such as TIMSS) use their own metrics and standards. For example, the TIMSS 2003 reports contain four international benchmarks: Advanced International Benchmark, High International Benchmark, Intermediate International Benchmark, and Low International Benchmark. However, these cut-scores are not as familiar to U.S. policymakers as the NAEP achievement levels. To interpret international results from TIMSS, using U.S. national benchmarks, this paper projects the NAEP achievement levels on to the TIMSS scale. This projection is accomplished through a secondary analysis of the linking study by Johnson and colleagues (2005).

Using projected NAEP achievement levels, the results of the four TIMSS surveys reported in this paper can be reinterpreted. In 1999 TIMSS mathematics, the number of counties with percentages of students significantly above the United States was: basic (16), proficient (11), and advanced (6). The number of counties with percentages of students significantly below the United States was: basic (16), proficient (17), and advanced (14). In 1999 TIMSS science, the number of counties with percentages of students significantly above the United States was: basic (6), proficient (2), and advanced (1). The number of counties with percentages of students significantly below the United States was: basic (14), proficient (16), and advanced (14).

Similarly, in 2003 TIMSS mathematics, the number of counties with percentages of students significantly above the United States was: basic (8), proficient (5), and advanced (5). The number of counties with percentages of students significantly below the United States was: basic (21), proficient (19), and advanced (16). In 2003 TIMSS science, the number of counties with percentages of students significantly above the United States was: basic (6), proficient (2), and advanced (1). The number of counties with percentages of students significantly below the United States was: basic (24), proficient (25), and advanced (17).

Looked at from the perspective of projected NAEP achievement levels, TIMSS results are more understandable. For example, tables 6, 8, 11, and 13 might be used to indicate which nations have world class educational achievement in mathematics or science. If a nation's average performance is at the proficient level, then it indicates that the typical student in that country is reaching a level of performance that meets U.S. standards. Interpreted this way, we find that the United States is a nation that is not meeting its own expectations.

The number of countries with averages at the various projected achievement levels is as follows. In 1999 TIMSS mathematics: below basic (12), basic (20), and proficient (6). In 1999 TIMSS science: below basic (18), basic (18), and proficient (2). In 2003 TIMSS mathematics: below basic (19), basic (22), and proficient (5). In 2003 TIMSS science: below basic (24), basic (20), and proficient (2). The United States average was at the basic level in all four surveys.

Overall, this report shows that interpreting international results in the light of U.S. standards can help make international patterns more visible to a U.S. audience—in particular, the outstanding educational achievements of several Asian countries, the mediocre performance of most English speaking and European countries, and the disturbingly low performance of many Middle Eastern and African nations.

One recommendation resulting from this study is that future international assessments should always include a linking study within the United States so that U.S. analysts and policymakers can better relate international results to national results. Future research might attempt to find methods to do the linking in ways that are simple and cost-effective. Furthermore, linking studies and validation studies in countries outside the United States would be an important contribution to testing the limits of linking methodology.

Technical Appendix

Section A: Error Variance Estimation

The linking procedure described in this paper is straightforward and easy to accomplish. The intermediate calculations of the error variance, however, are complex and tedious. This appendix describes the details of how the error variances reported in the paper were determined. Most of these analyses, especially those involving plausible values, were done as part of the study by Johnson et al. (2005). Furthermore, the analyses of plausible values have been well documented in the various technical manuals of both NAEP and TIMSS.

With statistical moderation, the estimated $\hat{TIMSS}_{level}$ is a linear transformation of $NAEP_{level}$. Therefore, the error variance in $\hat{TIMSS}_{level}$ is

$$\hat{\sigma}^2_{\hat{TIMSS}_{level}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{level}} + \hat{\sigma}^2_A + 2\left(NAEP_{level}\right)\hat{\sigma}_{AB} + \left(NAEP_{level}\right)^2 \hat{\sigma}^2_B. \qquad (1.3)$$

According to Johnson et al. (2005), the error variances of the parameters of the linear transformation, $\hat{\sigma}^2_A$, $2\hat{\sigma}^2_{AB}$ and $\hat{\sigma}^2_B$ can be approximated by Taylor-series linearization (Wolter, 1985)

$$\begin{aligned}
\hat{\sigma}^2_A &= \hat{B}^2\hat{\sigma}^2_{\mu_{NAEP}} + \hat{\sigma}^2_{\mu_{TIMSS}} + \hat{\mu}^2_{NAEP}\hat{B}^2\left[\frac{Var\left(\hat{\sigma}_{TIMSS}\right)}{\hat{\sigma}^2_{TIMSS}} + \frac{Var\left(\hat{\sigma}_{NAEP}\right)}{\hat{\sigma}^2_{NAEP}}\right] \\
2\hat{\sigma}_{AB} &= -2\hat{\mu}_{NAEP}\hat{B}^2\left[\frac{Var\left(\hat{\sigma}_{TIMSS}\right)}{\hat{\sigma}^2_{TIMSS}} + \frac{Var\left(\hat{\sigma}_{NAEP}\right)}{\hat{\sigma}^2_{NAEP}}\right] \qquad (1.4) \\
\hat{\sigma}^2_B &= \hat{B}^2\left[\frac{Var\left(\hat{\sigma}_{TIMSS}\right)}{\hat{\sigma}^2_{TIMSS}} + \frac{Var\left(\hat{\sigma}_{NAEP}\right)}{\hat{\sigma}^2_{NAEP}}\right].
\end{aligned}$$

In this particular application, we can treat the NAEP achievement levels as fixed, so there is no error associated with $NAEP_{level}$, therefore $\hat{B}^2\hat{\sigma}^2_{NAEP_{level}} = 0$. Equations (1.3) and(1.4), along with the data provided by Johnson et al. (2005), were used to derive the estimates in this paper.[4] The estimated achievement levels (along with their linking errors) are presented in table 3 for TIMSS mathematics and table 4 for TIMSS science. The standard error of linking reported in table 3 and table 4 is the square root of equation (1.3). The intermediate calculations for equations (1.3) and (1.4) are presented below.

Parameter estimates of the mean and standard deviation

The process begins with the analysis of plausible values for both NAEP and TIMSS. In both NAEP and TIMSS, five plausible values are used to represent the student's posterior distribution. Let us label the parameter we are estimating as "*t,*" and the number of plausible values as "*M,*" and the estimates of

[4] I wish to thank Tao Jiang at the American Institutes for Research® for providing the results of the analysis of plausible values for both NAEP and TIMSS from the study (Johnson et al. 2005) that allowed for the calculation of standard errors in this paper.

t as $\hat{t}_m$, for $m = 1, 2, ..M$. The average of the statistics is $t*$, where $t* = \sum_{m=1}^{M} \frac{t_m}{M}$. Tables 14A and 14B are the calculations for the parameter estimates of the means and standard deviations (SD).

Table 14A Estimating the mean and standard deviation in U.S. national samples (public schools) for grade 8 mathematics

	Plausible value 1	Plausible value 2	Plausible value 3	Plausible value 4	Plausible value 5	Mean plausible value (*t**)
2000 NAEP mathematics mean	274.505	274.467	274.329	274.297	274.480	274.416
1999 TIMSS mathematics mean	498.505	498.378	497.883	497.742	498.671	498.236
2000 NAEP mathematics SD	37.482	37.305	37.337	37.217	37.433	37.355
1999 TIMSS mathematics SD	86.481	88.451	89.410	89.047	88.549	88.388

Table 14B Estimating the mean and standard deviation in U.S. national samples (public schools) for grade 8 science

	Plausible value 1	Plausible value 2	Plausible value 3	Plausible value 4	Plausible value 5	Mean plausible value (*t**)
2000 NAEP science mean	149.301	149.229	148.998	149.037	149.382	149.189
1999 TIMSS science mean	509.305	510.657	510.460	509.437	512.086	510.389
2000 NAEP science SD	36.212	36.354	36.020	36.173	36.354	36.222
1999 TIMSS science SD	97.490	98.647	96.803	98.276	98.643	97.972

Sampling error variance of the mean and standard deviation

The error variances for the parameter estimates in tables 14A and 14B each have two components—error variance due to sampling ($U*$) and error variance due to measurement ($B*$). The sampling error in the estimates of the means and standard deviations were obtained by using a jackknife error variance approach for complex samples. The jackknife procedure was carried out for each plausible value and then averaged across all five plausible values. In the jackknife procedure, one primary sampling unit (PSU) is excluded; the sampling weights are redistributed across the other units within the stratum in which the PSU was excluded; the mean and standard deviation are calculated on the remaining PSUs; and the process is repeated until all PSUs have been excluded. After the jackknife procedure is carried out on each plausible value, the average across plausible values is $U* = \sum_{m=1}^{M} \frac{U_m}{M}$.

This process resulted in the variance estimates reported in tables 15A and 15B which are estimates of error variance due to sampling for the means and standard deviations.

Table 15A Sampling error variance of the mean and standard deviation (U^*) for grade 8 mathematics

Variance of NAEP mean 2000 mathematics from jackknife	0.640
Variance of TIMSS mean 1999 mathematics from jackknife	18.490
Variance of NAEP SD 2000 mathematics from jackknife	0.250
Variance of TIMSS SD 1999 mathematics from jackknife	6.250

Table 15B Sampling error variance of the mean and standard deviation (U^*) for grade 8 science

Variance of NAEP mean 2000 science from jackknife	0.490
Variance of TIMSS mean 1999 science from jackknife	25.000
Variance of NAEP SD 2000 science from jackknife	0.250
Variance of TIMSS SD 1999 science from jackknife	4.410

Measurement error variance of the mean and standard deviation

The error variance due to measurement is estimated by the variance between plausible values. This is estimated by $B^* = \frac{1+(1/M)}{M-1}\sum_{m=1}^{M}(t_m - t^*)^2$. The error variance due to measurement is in tables 16A and 16B.

Table 16A Measurement error variance of the mean and standard deviation (B^*) for grade 8 mathematics

Variance of NAEP mean 2000 mathematics from plausible values	0.011
Variance of TIMSS mean 1999 mathematics from plausible values	0.195
Variance of NAEP SD 2000 mathematics from plausible values	0.013
Variance of TIMSS SD 1999 mathematics from plausible values	1.544

Table 16B Measurement error variance of the mean and standard deviation (B^*) for grade 8 science

Variance of NAEP mean 2000 science from plausible values	0.033
Variance of TIMSS mean 1999 science from plausible values	1.511
Variance of NAEP SD 2000 science from plausible values	0.023
Variance of TIMSS SD 1999 science from plausible values	0.779

Total error variance of the mean and standard deviation

The total error variance is $V^* = U^* + B^*$ and is contained in tables 17A and 17B.

Table 17A Total error variance of the mean and standard deviation (V^*) for grade 8 mathematics

Variance of NAEP mean 2000 mathematics	0.651
Variance of TIMSS mean 1999 mathematics	18.685
Variance of NAEP SD 2000 mathematics	0.263
Variance of TIMSS SD 1999 mathematics	7.794

Table 17B Total error variance of the mean and standard deviation (V^*) for grade 8 science

Variance of NAEP mean 2000 science	0.523
Variance of TIMSS mean 1999 science	26.511
Variance of NAEP SD 2000 science	0.273
Variance of TIMSS SD 1999 science	5.189

Parameter estimates of the linking parameters *A* and *B*

The linking parameters are then calculated for each plausible value, using equation (1.2). The linking parameter estimates are then averaged over the five plausible values as reported in tables 18A and 18B.

Table 18A Estimating the linking parameters A and B in the U.S. national samples (public schools) for grade 8 mathematics

	Plausible value 1	Plausible value 2	Plausible value 3	Plausible value 4	Plausible value 5	Mean plausible value (*t**)
$\hat{A}$	–134.854	–152.393	–159.041	–158.554	–150.619	–151.077
$\hat{B}$	2.307	2.371	2.395	2.393	2.366	2.366

Table 18B Estimating the linking parameters A and B in the U.S. national samples (public schools) for grade 8 science

	Plausible value 1	Plausible value 2	Plausible value 3	Plausible value 4	Plausible value 5	Mean plausible value (*t**)
$\hat{A}$	107.351	105.720	110.029	104.531	106.752	106.877
$\hat{B}$	2.692	2.714	2.688	2.717	2.713	2.705

Sampling error variance of the linking parameters *A* and *B*

The error variance of the linking parameters estimates $\hat{A}$ and $\hat{B}$ is found by equation (1.4). The linking error variance also has two components—one due to sampling and one due to measurement error. The quantities needed to estimate the error variance in the linking parameters due to sampling are contained in tables 16A and 16B. The quantities needed to estimate the error variance in the linking parameters due to measurement error are contained in tables 17A and 17B. Substituting the estimates in tables 16A and 16B in equation (1.4), we have the error variance in the linking parameters due to sampling. These are reported in tables 19A and 19B.

Table 19A Sampling error variance in NAEP–TIMSS linking parameters for mathematics

Error variance in A, $\left(\hat{\sigma}^2_{A(s)}\right)$	434.901
Two times the covariance between A and B, $2\left(\hat{\sigma}_{AB(s)}\right)$	–3.009
Error variance in B, $\left(\hat{\sigma}_{B(s)}\right)$	0.005

Table 19B Sampling error variance in NAEP–TIMSS linking parameters for science

Error variance in A, $\left(\hat{\sigma}^2_{A(s)}\right)$	108.740
Two times the covariance between A and B, $2\left(\hat{\sigma}_{AB(s)}\right)$	–1.086
Error variance in B, $\left(\hat{\sigma}_{B(s)}\right)$	0.004

Measurement error variance of the linking parameters *A* and *B*

Substituting the estimates in tables 17A and 17B in equation (1.4) provides the error variance in the linking parameters due to measurement error, as reported in tables 20A and 20B.

Table 20A Measurement error variance in NAEP–TIMSS linking parameters for grade 8 mathematics

Error variance in A, $\left(\hat{\sigma}^2_{A(m)}\right)$	87.575
Two times the covariance between A and B, $2\left(\hat{\sigma}_{AB(m)}\right)$	–0.636
Error variance in B, $\left(\hat{\sigma}_{B}\right)$	0.001

Table 20B Measurement error variance in NAEP–TIMSS linking parameters for grade 8 science

Error variance in A, $\left(\hat{\sigma}^2_{A(m)}\right)$	14.040
Two times the covariance between A & B, $2\left(\hat{\sigma}_{AB(m)}\right)$	–0.165
Error variance in B, $\left(\hat{\sigma}_{B(m)}\right)$	0.001

Total error variance of the linking parameters *A* and *B*

The sum of the sampling error variances in tables 19A and 19B and the measurement error variances in tables 20A and 20B yield the total error variances in the linking parameters reported in tables 21A and 21B.

Table 21A Total error variance in NAEP–TIMSS linking parameters for grade 8 mathematics

Error variance in A, $\left(\hat{\sigma}^2_{A}\right)$	522.476
Two times the covariance between A and B, $2\left(\hat{\sigma}_{AB}\right)$	–3.645
Error variance in B, $\left(\hat{\sigma}_{B}\right)$	0.007

Table 21B Total error variance in NAEP–TIMSS linking parameters for grade 8 science

Error variance in A, $\left(\hat{\sigma}^2_{A}\right)$	122.781
Two times the covariance between A and B, $2\left(\hat{\sigma}_{AB}\right)$	–1.251
Error variance in B, $\left(\hat{\sigma}_{B}\right)$	0.004

Linking error variance (due to sampling) of the projected NAEP achievement levels

The linking error variance of the projected NAEP achievement levels on the TIMSS scale is found in equation (1.3). The linking error variance also has two components—one due to sampling, and one due to measurement error. The quantities needed to estimate the error variance in the projected achievement levels due to sampling are contained in tables 19A and 19B. The quantities needed to estimate the error variance in the linking parameters due to measurement error are contained in tables 20A and 20B.

Substituting the estimates in tables 19A and 19B in equation (1.3), we have the linking error variance in the projected achievement levels due to sampling. These are reported in tables 22A and 22B.[5]

Table 22A Error variance in linking due to sampling for NAEP achievement levels projected onto TIMSS grade 8 mathematics scale

$\hat{\sigma}^2_{TIMSS_{basic}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{basic}} + \hat{\sigma}^2_{A(s)} + 2\left(NAEP_{basic}\right)\hat{\sigma}_{AB(s)} + \left(NAEP_{basic}\right)^2\hat{\sigma}^2_{B(s)}$	22.918
$\hat{\sigma}^2_{TIMSS_{prof}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{prof}} + \hat{\sigma}^2_{A(s)} + 2\left(NAEP_{prof}\right)\hat{\sigma}_{AB(s)} + \left(NAEP_{prof}\right)^2\hat{\sigma}^2_{B(s)}$	25.387
$\hat{\sigma}_{TIMSS_{adv}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{adv}} + \hat{\sigma}^2_{A(s)} + 2\left(NAEP_{adv}\right)\hat{\sigma}_{AB(s)} + \left(NAEP_{adv}\right)^2\hat{\sigma}^2_{B(s)}$	40.889

Table 22B Error variance in linking due to sampling for NAEP achievement levels projected onto TIMSS grade 8 science scale

$\hat{\sigma}^2_{TIMSS_{basic}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{basic}} + \hat{\sigma}^2_{A(s)} + 2\left(NAEP_{basic}\right)\hat{\sigma}_{AB(s)} + \left(NAEP_{basic}\right)^2\hat{\sigma}^2_{B(s)}$	27.883
$\hat{\sigma}^2_{TIMSS_{prof}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{prof}} + \hat{\sigma}^2_{A(s)} + 2\left(NAEP_{prof}\right)\hat{\sigma}_{AB(s)} + \left(NAEP_{prof}\right)^2\hat{\sigma}^2_{B(s)}$	29.319
$\hat{\sigma}_{TIMSS_{adv}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{adv}} + \hat{\sigma}^2_{A(s)} + 2\left(NAEP_{adv}\right)\hat{\sigma}_{AB(s)} + \left(NAEP_{adv}\right)^2\hat{\sigma}^2_{B(s)}$	40.330

Linking error variance (due to measurement) of the projected NAEP achievement levels

Substituting the estimates in tables 20A and 20B in equation (1.3) provides the linking error variance in the projected achievement levels due to measurement error as reported in tables 23A and 23B.

Table 23A Error variance in linking due to measurement for NAEP achievement levels projected onto TIMSS grade 8 mathematics scale

$\hat{\sigma}^2_{TIMSS_{basic}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{basic}} + \hat{\sigma}^2_{A(m)} + 2\left(NAEP_{basic}\right)\hat{\sigma}_{AB(m)} + \left(NAEP_{basic}\right)^2\hat{\sigma}^2_{B(m)}$	0.435
$\hat{\sigma}^2_{TIMSS_{prof}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{prof}} + \hat{\sigma}^2_{A(m)} + 2\left(NAEP_{prof}\right)\hat{\sigma}_{AB(m)} + \left(NAEP_{prof}\right)^2\hat{\sigma}^2_{B(m)}$	0.957
$\hat{\sigma}_{TIMSS_{adv}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{adv}} + \hat{\sigma}^2_{A(m)} + 2\left(NAEP_{adv}\right)\hat{\sigma}_{AB(m)} + \left(NAEP_{adv}\right)^2\hat{\sigma}^2_{B(m)}$	4.236

[5] Since the NAEP achievement levels are a known parameter, we assume throughout this paper that $\hat{B}^2\hat{\sigma}^2_{NAEP_{ach\ level}}$ is equal to zero.

Table 23B Error variance in linking due to measurement for NAEP achievement levels projected onto TIMSS grade 8 science scale

$\hat{\sigma}^2_{TIMSS_{basic}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{basic}} + \hat{\sigma}^2_{A(m)} + 2\left(NAEP_{basic}\right)\hat{\sigma}_{AB(m)} + \left(NAEP_{basic}\right)^2\hat{\sigma}^2_{B(m)}$	1.719
$\hat{\sigma}^2_{TIMSS_{prof}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{prof}} + \hat{\sigma}^2_{A(m)} + 2\left(NAEP_{prof}\right)\hat{\sigma}_{AB(m)} + \left(NAEP_{prof}\right)^2\hat{\sigma}^2_{B(m)}$	1.938
$\hat{\sigma}_{TIMSS_{adv}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{adv}} + \hat{\sigma}^2_{A(m)} + 2\left(NAEP_{adv}\right)\hat{\sigma}_{AB(m)} + \left(NAEP_{adv}\right)^2\hat{\sigma}^2_{B(m)}$	3.616

Total linking error variance of the projected NAEP achievement levels

The sum of the linking error variance due to sampling in tables 22A and 22B and the linking error variance due to measurement tables 23A and 23B yields the total linking error variances in the projected achievement levels on the TIMSS scale reported in tables 24A and 24B.

Table 24A Total error variance in linking for NAEP achievement levels projected onto TIMSS grade 8 mathematics scale

$\hat{\sigma}^2_{TIMSS_{basic}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{basic}} + \hat{\sigma}^2_{A} + 2\left(NAEP_{basic}\right)\hat{\sigma}_{AB} + \left(NAEP_{basic}\right)^2\hat{\sigma}^2_{B}$	23.353
$\hat{\sigma}^2_{TIMSS_{prof}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{prof}} + \hat{\sigma}^2_{A} + 2\left(NAEP_{prof}\right)\hat{\sigma}_{AB} + \left(NAEP_{prof}\right)^2\hat{\sigma}^2_{B}$	26.343
$\hat{\sigma}_{TIMSS_{adv}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{adv}} + \hat{\sigma}^2_{A} + 2\left(NAEP_{adv}\right)\hat{\sigma}_{AB} + \left(NAEP_{adv}\right)^2\hat{\sigma}^2_{B}$	45.124

Table 24B Total error variance in linking for NAEP achievement levels projected onto TIMSS grade 8 science scale

$\hat{\sigma}^2_{TIMSS_{basic}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{basic}} + \hat{\sigma}^2_{A} + 2\left(NAEP_{basic}\right)\hat{\sigma}_{AB} + \left(NAEP_{basic}\right)^2\hat{\sigma}^2_{B}$	29.602
$\hat{\sigma}^2_{TIMSS_{prof}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{prof}} + \hat{\sigma}^2_{A} + 2\left(NAEP_{prof}\right)\hat{\sigma}_{AB} + \left(NAEP_{prof}\right)^2\hat{\sigma}^2_{B}$	31.257
$\hat{\sigma}_{TIMSS_{adv}} = \hat{B}^2\hat{\sigma}^2_{NAEP_{adv}} + \hat{\sigma}^2_{A} + 2\left(NAEP_{adv}\right)\hat{\sigma}_{AB} + \left(NAEP_{adv}\right)^2\hat{\sigma}^2_{B}$	43.946

The standard errors of linking reported in tables 3 and 4 are the square roots of the linking error variances in tables 24A and 24B.

It is instructive to compare the standard error of linking for the projected NAEP mean to the standard error of linking for the projected NAEP achievement levels. Because the linking error is smaller at the mean, the standard error of linking for the NAEP projected achievement levels should be larger than for the mean. In fact, this is the case. The standard error of linking for the projected mean of 498 in mathematics is 4.73 and for the projected mean of 510 in science is 5.43. In both cases, the standard error of linking for the mean is smaller than the standard error of linking for the achievement levels reported in tables 3 and 4.

One interesting question in linking studies is, "How much of the linking error is due to sampling and how much is due to test unreliability (or measurement error)?" In this study, we can answer that question by comparing the error variances in tables 22A and 22B, and 23A and 23B, to 24A and 24B. Tables 24A and 24B show the percent of error variance accounted for by sampling and measurement error.

Table 25A Variance components of linking error for NAEP achievement levels projected on to the TIMSS grade 8 mathematics scale

	Sampling	Measurement
Basic	98.1%	1.9%
Proficient	96.4%	3.6%
Advanced	90.6%	9.4%

Table 25B Variance components of linking error for NAEP achievement levels projected on to the TIMSS grade 8 science scale

	Sampling	Measurement
Basic	94.2%	5.8%
Proficient	93.8%	6.2%
Advanced	91.8%	8.2%

The main message of tables 25A and 25B is that the vast majority of linking error is due to sampling. However, measurement error becomes a larger percentage of the linking error in the tails of the achievement distribution. This is why the measurement error for the advanced achievement level is a larger component of the linking error variance. The advanced achievement level is very high on the scale, where the measurement error is larger.

Linking error variance for the percent at and above projected achievement levels

So far in this technical appendix, all the error variances have been calculated in the scale score metric. However, the report is really about the percentages of students at and above various achievement levels (inverse cumulative percentages). Thus we must express the standard errors of linking in the inverse cumulative percentage metric as well as the scale score metric. This was done by making the assumption that the population distribution in each country is approximately normal. We know this assumption may not be true in some very low-performing and very high-performing countries. However, even in these circumstances, the normality assumption should still provide reasonable approximations. Suppose that the TIMSS achievement of students θ is normally distributed in country j with $\theta \sim N\left(\mu_j, \sigma_j\right)$. Estimates, $\hat{\mu}_j$ and $\hat{\sigma}_j$ of μ_j and σ_j are available from the published international reports of 1999 TIMSS and 2003 TIMSS. Let θ_C represent the cut-score on the TIMSS scale for the projected NAEP achievement level. Given the normality assumption, the percentage of students at and above each projected achievement level is

$$P_j(\theta > \theta_C) = \left[1 - \Phi\left(\frac{\theta_C - \hat{\mu}_j}{\hat{\sigma}_j}\right)\right] * 100, \tag{1.5}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of a standard normal distribution.

However, we know that there is linking error (*LE*) in the projected achievement levels. Let $\theta_{C+\sigma_{LE}}$ be the upper limit of the margin of error interval for linking and $\theta_{C-\sigma_{LE}}$ be the lower limit. Then the percentage, P_j of students at and above the achievement level θ_C is between the upper and lower limit of the margin of error interval. The upper and lower limits are

$$\begin{aligned} P_{j+LE}(\theta > \theta_{C+\sigma_{LE}}) &= \left[1 - \Phi\left(\frac{\theta_{C+\sigma_{LE}} - \hat{\mu}_j}{\hat{\sigma}_j}\right)\right] * 100, \text{ and} \\ P_{j-LE}(\theta > \theta_{C-\sigma_{LE}}) &= \left[1 - \Phi\left(\frac{\theta_{C-\sigma_{LE}} - \hat{\mu}_j}{\hat{\sigma}_j}\right)\right] * 100, \end{aligned} \tag{1.6}$$

Although the upper and lower limits of the margin of error (P_{j+LE} and P_{j-LE}) are asymmetrical around P_j, a rough standard error of linking in the inverse cumulative percent metric can be obtained by

$$\sigma_{LEj} = \frac{P_{j-LE} - P_{j+LE}}{2} \tag{1.7}$$

Sampling error variance for the percent at and above projected achievement levels

Because TIMSS is a survey that is administered in each country, all statistics derived from it will have sampling error. Therefore, the percent of students at and above each projected achievement level P_j will have sampling error associated with it in equation (1.5). The sampling error can be estimated from the published international reports by calculating the standard error of a percentage

$$\sigma_{SEj} = \sqrt{\frac{P_j\left(1 - P_j\right)}{eff\,(n_j)}}\,. \tag{1.8}$$

The quantity $eff\,(n_j)$ is the effective sample size (i.e., the actual sample size of the survey divided by the design effect). The effective sample size can be determined from the published reports of the survey if we know the standard deviation of scaled scores, SD_j, and the standard error of the mean of scaled scores, SEM_j, (both of which are reported in the international publications) by the formula

$$eff\,(n_j) = \left(\frac{SD_j}{SEM_j}\right)^2. \tag{1.9}$$

Total error variance for the percent at and above projected achievement levels

The total standard error for the percent of student at and above each achievement level P_j is the square root of the sum of the squared linking error (1.7) and squared sampling error (1.8).

$$\sigma_{Ej} = \sqrt{\sigma^2_{LEj} + \sigma^2_{SEj}} \tag{1.10}$$

These margins of error are reported in tables 5, 7, 10, and 12.

Section B: Linking

Mislevy (1992) and Linn (1993) have described many of the conceptual and statistical issues associated with linking assessments. They have outlined four forms of statistical linking: equating, calibration, projection, and statistical moderation. A further explication of the differences is provided here.

The three assumptions that distinguish the different forms of statistical linking are that two tests (call them *X* and *Y*) have true scores that are highly correlated, measure the same content, and are equally reliable. These assumptions are displayed in table 30.

Table 30 Statistically linking test *X* and test *Y*

	Equating	Calibration	Projection	Moderation
High true score correlation	x[6]	x[6]	x	
Same content	x	x		
Equal reliability	x			

In *equating,* both tests, *X* and *Y*, have been designed and developed to be equally reliable, and each measures the same content. Equating is used when the goal is to relate two alternate forms of the same test, such as alternate forms of the ACT or the SAT. Under these conditions, the only difference between the two tests is the metric, such as expressing temperature in terms of Fahrenheit or Celsius. In equating the distributions of test *X* and *Y* are aligned or matched up directly. The matching can be done with equipercentile equating or linear equating, and the distributions can be either observed score distributions or estimates of the true score distributions. When the three assumptions (high correlation, same content, and equal reliability) are met:

- the linking function should be the same for *X* expressed in terms of *Y*, and for *Y* expressed in terms of *X*, and
- the linking function should be the same for different subgroups, across contexts and time.

In *calibration* (for example with the use of item-response theory), two tests are assumed to measure the same content, but they are not equally reliable. For example, one test *X* might be a long test whereas the other test *Y* is short. The two versions of the test are not equated, but they are indirectly comparable because they have been calibrated to a common scale θ . This type of linking is done across grades and across years in NAEP, TIMSS, most state criterion-referenced tests, and most nationally standardized norm-referenced tests. Calibration procedures provide unbiased estimates for individual students and means, but additional statistical machinery is needed to accurately estimate group characteristics such as

[6] The true-score correlation between *X* and *Y* is assumed to equal 1.0.

the variance or the percent at and above achievement levels. When the two assumptions (high correlation and same content) are met:

- the linking function between *X* and θ (e.g., the test characteristic curve) is different from the linking function between *Y* and θ,
- both *X* and *Y* can be used to get unbiased estimates of θ for individual students (although the error in the estimates will be higher for *Y*), however
- the observed score distributions of *X* for groups do not match the observed score distributions for *Y*.

In *projection,* a regression equation uses the correlation between the two tests to predict the scores on one test *Y* from those of another test *X*. There is no assumption that the two tests measure the same content or that they are equally reliable. With projection, there is no longer a symmetric relationship between one test and the other. The conversion table for predicting the first test from the second is different from the table predicting the second test from the first. When the assumption of high correlation is met:

- the linking function for *X* expressed in terms of *Y* (e.g., regression equation) will be different from the linking function for *Y* expressed in terms of *X*, and
- the linking function will likely be different for different subgroups, across contexts and time.

In *statistical moderation*, the scores on the first test *X* are adjusted to have the same distributional characteristics as the scores on the second test *Y*. In this case *X* is linked to *Y*. This is typically done by matching the means and standard deviations of *X* and *Y*, or matching their percentile ranks. The usual assumption is that both, *X* and *Y,* have been administered to comparable populations of students (e.g., the student populations taking both tests are randomly equivalent). Statistical moderation typically does not use the correlation between the two tests. When statistical moderation is used:

- the linking function for *X* expressed in terms of *Y* (e.g., a z-score equivalency) will be different from the linking function for *Y* expressed in terms of *X*,
- the linking function will likely be different for different subgroups, across contexts and time, and
- the degree of the relationship between *X* and *Y* is typically unknown.

Section C: Additional Significance Testing

Simple comparisons versus multiple comparisons

All of the significance tests performed in tables 5, 7, 10, and 12 are simple comparisons. This means the percent at and above each projected achievement level in each country is compared to that of the United States. If we refer to the United States as A and any other country as B, then the 95% confidence interval is

$$95\%CI = \pm Z_{\alpha/2}\sqrt{\sigma^2_{E(A)} + \sigma^2_{E(B)}}\ . \tag{1.11}$$

The confidence interval is strictly true only if we compare one country to the United States. If we compare many countries to the United States, then the overall confidence interval is smaller. In 1999, TIMSS used a Bonferroni adjustment to the alpha level to keep the overall alpha level equal to 0.05 and the overall confidence interval at 95%. In the 2003 TIMSS, this practice was discontinued. If the reader wishes to make the Bonferroni adjustment, it would be done as follows. If there are k countries in the study, then we can make $k-1$ comparisons to the United States for each projected achievement level. In the 1999 TIMSS, $k = 38$; and in the 2003 TIMSS, $k = 46$. The alpha level is therefore divided by $k-1$. Each comparison is made with an alpha $\alpha/(k-1)$. To make $k-1$ multiple comparisons to the United States and keep the overall confidence interval at 95%, this can be done by using equation (1.11) with

$$95\%CI = \pm Z_{\alpha/2(k-1)}\sqrt{\sigma^2_{E(A)} + \sigma^2_{E(B)}}\ .$$

Additional Significance Tests

Tables 5, 7, 10, and 12 compare each country to the United States. For example, in table 10 there are $k = 46$ countries, so there are $k(k-1)/2 = 1035$ possible comparisons. Only $k-1 = 45$ of the 1,035 possible comparisons are presented in table 10 (those that involve the United States). If the reader wishes to select another country (e.g., Canada) and compare every other country to the selected country, tables 26, 27, 28, and 29 can be used for the projected proficient achievement level.

Table 26 Comparisons for 1999 TIMSS in mathematics with each country compared to another country for the percent estimated to be proficient based on NAEP achievement levels projected on to the TIMSS scale

Country	Singapore	Korea, Rep. of	Hong Kong, SAR	Japan	Chinese Taipei	Belgium (Flemish)	Netherlands	Hungary	Slovak Republic	Slovenia	Canada	Russian Federation	Australia	Czech Republic	Malaysia	Bulgaria	Finland	**United States**	Latvia (LSS)	England	New Zealand	Italy	Romania	Israel	Lithuania	Cyprus	Moldova	Thailand	Macedonia, Rep. of	Jordan	Turkey	Indonesia	Iran, Islamic Rep. of	Tunisia	Chile	Philippines	Morocco	South Africa
Singapore				▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Korea, Rep. of						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hong Kong, SAR						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Japan	▼					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Chinese Taipei	▼					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Belgium (Flemish)	▼	▼	▼	▼	▼			▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Netherlands	▼	▼	▼	▼	▼													▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hungary	▼	▼	▼	▼	▼	▼											▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovak Republic	▼	▼	▼	▼	▼	▼												▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovenia	▼	▼	▼	▼	▼	▼											▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Canada	▼	▼	▼	▼	▼	▼												▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Russian Federation	▼	▼	▼	▼	▼	▼													▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Australia	▼	▼	▼	▼	▼	▼													▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Czech Republic	▼	▼	▼	▼	▼	▼														▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Malaysia	▼	▼	▼	▼	▼	▼														▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Bulgaria	▼	▼	▼	▼	▼	▼																▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Finland	▼	▼	▼	▼	▼	▼		▼		▼												▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
United States	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Latvia (LSS)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
England	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
New Zealand	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼												▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Italy	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Romania	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼												▲	▲	▲	▲	▲	▲	▲	▲	▲
Israel	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											▲	▲	▲	▲	▲	▲	▲	▲	▲
Lithuania	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											▲	▲	▲	▲	▲	▲	▲	▲	▲
Cyprus	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼										▲	▲	▲	▲	▲	▲	▲	▲	▲
Moldova	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼										▲	▲	▲	▲	▲	▲	▲	▲
Thailand	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼										▲	▲	▲	▲	▲	▲	▲	▲
Macedonia, Rep. of	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼									▲	▲	▲	▲	▲	▲	▲	▲
Jordan	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼						▲	▲	▲	▲	▲	▲	▲
Turkey	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼						▲	▲	▲	▲
Indonesia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼					▲	▲	▲	▲
Iran, Islamic Rep. of	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼						▲	▲	▲
Tunisia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼						▲	▲	▲
Chile	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼						▲
Philippines	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼				
Morocco	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼				
South Africa	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼			

Select a country on the left, then read across the row for comparisons with all other countries listed above. The symbol ▲ indicates the percent estimated to be proficient for the country on the left is significantly higher than the comparison country above. The symbol ▼ indicates the percent estimated to be proficient for the country on the left is significantly lower than the comparison country above. With a 95% confidence interval, 5% of the comparisons will be significant by chance.

Table 27 Comparisons for 1999 TIMSS in science with each country compared to another country for the percent estimated to be proficient based on NAEP achievement levels projected on to the TIMSS scale

Country	Chinese Taipei	Singapore	Hungary	Korea, Rep. of	Japan	Netherlands	Australia	England	Czech Republic	Slovenia	Russian Federation	Finland	Slovak Republic	Canada	Belgium (Flemish)	Bulgaria	Hong Kong, SAR	**United States**	New Zealand	Latvia (LSS)	Italy	Malaysia	Israel	Lithuania	Romania	Macedonia, Rep. of	Jordan	Moldova	Thailand	Cyprus	Iran, Islamic Rep. of	Indonesia	Chile	Turkey	Philippines	Tunisia	Morocco	South Africa
Chinese Taipei										▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Singapore										▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hungary																			▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Korea, Rep. of																			▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Japan																				▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Netherlands																				▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Australia																				▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
England																				▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Czech Republic																				▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovenia	▼	▼																		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Russian Federation	▼	▼																			▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Finland	▼	▼																			▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovak Republic	▼	▼																			▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Canada	▼	▼																			▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Belgium (Flemish)	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Bulgaria	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hong Kong, SAR	▼	▼																								▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
United States	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
New Zealand	▼	▼	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Latvia (LSS)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																				▲	▲	▲	▲	▲	▲	▲	▲	▲
Italy	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																▲	▲	▲	▲	▲	▲	▲	▲	▲
Malaysia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																	▲	▲	▲	▲	▲	▲	▲	▲
Israel	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		▼													▲	▲	▲	▲	▲	▲	▲	▲
Lithuania	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		▼													▲	▲	▲	▲	▲	▲	▲	▲
Romania	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		▼													▲	▲	▲	▲	▲	▲	▲	▲
Macedonia, Rep. of	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼													▲	▲	▲	▲	▲	▲	▲
Jordan	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼													▲	▲	▲	▲	▲	▲	▲
Moldova	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼													▲	▲	▲	▲	▲	▲	▲
Thailand	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																▲	▲	▲	▲
Cyprus	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼														▲	▲	▲	▲
Iran, Islamic Rep. of	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											▲	▲	▲
Indonesia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼									▲	▲
Chile	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼									▲	▲
Turkey	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼									▲	▲
Philippines	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼								
Tunisia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼							
Morocco	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼				
South Africa	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼				

Select a country on the left, then read across the row for comparisons with all other countries listed above. The symbol ▲ indicates the percent estimated to be proficient for the country on the left is significantly higher than the comparison country above. The symbol ▼ indicates the percent estimated to be proficient for the country on the left is significantly lower than the comparison country above. With a 95% confidence interval, 5% of the comparisons will be significant by chance.

Table 28 Comparisons for 2003 TIMSS in mathematics with each country compared to another country for the percent estimated to be proficient based on NAEP achievement levels projected on to the TIMSS scale

Country	Singapore	Hong Kong, SAR	Korea, Rep. of	Chinese Taipei	Japan	Belgium (Flemish)	Netherlands	Hungary	Estonia	Slovak Republic	Australia	Russian Federation	Malaysia	**United States**	Latvia	Lithuania	Israel	England	Scotland	New Zealand	Sweden	Serbia	Slovenia	Romania	Armenia	Italy	Bulgaria	Moldova, Rep. of	Cyprus	Norway	Macedonia, Rep. of	Jordan	Egypt	Indonesia	Palestinian Nat'l Auth.	Lebanon	Iran, Islamic Rep. of	Chile	Bahrain	Philippines	Tunisia	Morocco	Botswana	South Africa	Saudi Arabia	Ghana
Singapore					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hong Kong, SAR						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Korea, Rep. of						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Chinese Taipei						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Japan	▼					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Belgium (Flemish)	▼	▼	▼	▼	▼											▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Netherlands	▼	▼	▼	▼	▼												▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hungary	▼	▼	▼	▼	▼													▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Estonia	▼	▼	▼	▼	▼																▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovak Republic	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Australia	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Russian Federation	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Malaysia	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
United States	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Latvia	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Lithuania	▼	▼	▼	▼	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Israel	▼	▼	▼	▼	▼	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
England	▼	▼	▼	▼	▼	▼	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Scotland	▼	▼	▼	▼	▼	▼	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
New Zealand	▼	▼	▼	▼	▼	▼	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Sweden	▼	▼	▼	▼	▼	▼	▼	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Serbia	▼	▼	▼	▼	▼	▼	▼	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovenia	▼	▼	▼	▼	▼	▼	▼	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Romania	▼	▼	▼	▼	▼	▼	▼	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Armenia	▼	▼	▼	▼	▼	▼	▼	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Italy	▼	▼	▼	▼	▼	▼	▼	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Bulgaria	▼	▼	▼	▼	▼	▼	▼	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Moldova, Rep. of	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																▲		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Cyprus	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼														▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Norway	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼										▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Macedonia, Rep. of	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼								▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Jordan	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼										▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Egypt	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼										▲		▲	▲	▲	▲	▲	▲
Indonesia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		▼												▲	▲	▲	▲	▲	▲
Palestinian Nat'l Auth.	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼										▲	▲	▲	▲	▲	▲
Lebanon	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼												▲	▲	▲	▲
Iran, Islamic Rep. of	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											▲	▲	▲	▲
Chile	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼												▲	▲	▲
Bahrain	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼										▲	▲	▲	▲

Country	Singapore	Hong Kong, SAR	Korea, Rep. of	Chinese Taipei	Japan	Belgium (Flemish)	Netherlands	Hungary	Estonia	Slovak Republic	Australia	Russian Federation	Malaysia	**United States**	Latvia	Lithuania	Israel	England	Scotland	New Zealand	Sweden	Serbia	Slovenia	Romania	Armenia	Italy	Bulgaria	Moldova, Rep. of	Cyprus	Norway	Macedonia, Rep. of	Jordan	Egypt	Indonesia	Palestinian Nat'l Auth.	Lebanon	Iran, Islamic Rep. of	Chile	Bahrain	Philippines	Tunisia	Morocco	Botswana	South Africa	Saudi Arabia	Ghana
Philippines	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼														
Tunisia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											
Morocco	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											
Botswana	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		▼							
South Africa	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼							
Saudi Arabia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼							
Ghana	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼							

Select a country on the left, then read across the row for comparisons with all other countries listed above. The symbol ▲ indicates the percent estimated to be proficient for the country on the left is significantly higher than the comparison country above. The symbol ▼ indicates the percent estimated to be proficient for the country on the left is significantly lower than the comparison country above. With a 95% confidence interval, 5% of the comparisons will be significant by chance.

Table 29 Comparisons for 2003 TIMSS in science with each country compared to another country for the percent estimated to be proficient based on NAEP achievement levels projected on to the TIMSS scale

Country	Singapore	Chinese Taipei	Korea, Rep. of	Hong Kong, SAR	Japan	Estonia	England	Hungary	**United States**	Netherlands	Australia	Sweden	New Zealand	Slovak Republic	Lithuania	Slovenia	Russian Federation	Scotland	Belgium (Flemish)	Latvia	Malaysia	Israel	Bulgaria	Italy	Jordan	Norway	Romania	Serbia	Macedonia, Rep. of	Moldova, Rep. of	Armenia	Egypt	Palestinian Nat'l Auth.	Iran, Islamic Rep. of	Cyprus	Bahrain	Chile	Indonesia	Philippines	Lebanon	Saudi Arabia	Botswana	South Africa	Morocco	Ghana	Tunisia
Singapore							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Chinese Taipei									▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Korea, Rep. of												▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hong Kong, SAR														▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Japan														▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Estonia																	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
England	▼																		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hungary	▼																		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
United States	▼	▼																				▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Netherlands	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Australia	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Sweden	▼	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
New Zealand	▼	▼	▼																									▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovak Republic	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Lithuania	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovenia	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Russian Federation	▼	▼	▼	▼	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Scotland	▼	▼	▼	▼	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Belgium (Flemish)	▼	▼	▼	▼	▼	▼	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Latvia	▼	▼	▼	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Malaysia	▼	▼	▼	▼	▼	▼	▼	▼																								▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Israel	▼	▼	▼	▼	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Bulgaria	▼	▼	▼	▼	▼	▼	▼	▼	▼																							▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Italy	▼	▼	▼	▼	▼	▼	▼	▼	▼		▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Jordan	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																					▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Norway	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Romania	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																						▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Serbia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Macedonia, Rep. of	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Moldova, Rep. of	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Armenia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼																	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Egypt	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼													▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Palestinian Nat'l Auth.	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼															▲	▲	▲	▲	▲	▲	▲
Iran, Islamic Rep. of	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼														▲	▲	▲	▲	▲	▲
Cyprus	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼													▲	▲	▲	▲	▲	▲
Bahrain	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼												▲	▲	▲	▲	▲	▲
Chile	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											▲	▲	▲	▲
Indonesia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼														▲

Country	Singapore	Chinese Taipei	Korea, Rep. of	Hong Kong, SAR	Japan	Estonia	England	Hungary	**United States**	Netherlands	Australia	Sweden	New Zealand	Slovak Republic	Lithuania	Slovenia	Russian Federation	Scotland	Belgium (Flemish)	Latvia	Malaysia	Israel	Bulgaria	Italy	Jordan	Norway	Romania	Serbia	Macedonia, Rep. of	Moldova, Rep. of	Armenia	Egypt	Palestinian Nat'l Auth.	Iran, Islamic Rep. of	Cyprus	Bahrain	Chile	Indonesia	Philippines	Lebanon	Saudi Arabia	Botswana	South Africa	Morocco	Ghana	Tunisia
Philippines	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼													▲	▲
Lebanon	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼											▲	▲	▲
Saudi Arabia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼										
Botswana	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼										
South Africa	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼									
Morocco	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼			▼						
Ghana	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		▼	▼						
Tunisia	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼						

Select a country on the left, then read across the row for comparisons with all other countries listed above. The symbol ▲ indicates the percent estimated to be proficient for the country on the left is significantly higher than the comparison country above. The symbol ▼ indicates the percent estimated to be proficient for the country on the left is significantly lower than the comparison country above. With a 95% confidence interval, 5% of the comparisons will be significant by chance.

References

Beaton, A. E., and E. J. Gonzales. *Comparing the NAEP Trial State Assessment Results with the IAEP International Results* (report prepared for the National Academy of Education Panel on the NAEP Trial State Assessment). Stanford, CA: National Academy of Education, 1993.

Braswell, J. S, A. D. Lutkus, W. S. Grigg, S. L. Santapau, B. S. Tay-Lim, and M. S. Johnson. *The Nations Report Card: Mathematics 2000.* Washington, DC: National Center for Educational Statistics, 2001.

Johnson, E. G., and A. Siengondorf. *Linking the National Assessment of Educational Progress and the Third International Mathematics and Science Study: Eighth Grade Results*. (Publication No. NCES 98-500). Washington, DC: National Center for Education Statistics, 1998.

Johnson, E. G., J. Cohen, W.-H. Chen, T. Jiang, and Y. Zhang. *2000 NAEP–1999 TIMSS Linking Report*. (Publication No. 2005-01). U.S. Department of Education. Washington DC: National Center for Education Statistics, 2005.

LaPointe, A. E., N. A. Mead, and G. W. Phillips. *A World of Differences: An International Assessment of Mathematics and Science*. (Report No. 19-CAEP-01). Princeton, NJ: Educational Testing Service, 1989.

LaPointe, A. E., N. A. Mead, and J. M. Askew. *Learning Mathematics.* Princeton, NJ: Educational Testing Service, 1992.

Linn, R. L. "Linking Results of District Assessments." *Applied Measurement in Education*, 6 (1993): 83–102.

Martin, M. O., Ina V. S. Mullis, E. J. Gonzalez, K. D. Gregory, T. A. Smith, S. J. Chrostowski, R. A. Garden, and K. M. O'Connor. *TIMSS 1999 International Science Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade.* Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College. 2000.

Martin, M. O., Ina V. S. Mullis, E. J. Gonzalez, and S. J. Chrostowski. *TIMSS 2003 International Science Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades,* Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College, 2004.

Mislevy, R. J. *Linking Educational Assessments: Concepts, Issues, Methods and Prospects*. Princeton, NJ: Policy Information Center, Educational Testing Service, 1992.

Mullis, Ina V.S., M. O. Martin, E. J. Gonzales, K. D. Gregory, R. A. Garden, K. M. O'Connor, S. J. Chrostowski, and T. A. Smith. *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of The Third International Mathematics and Science Study at the Eighth Grade*, Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College, 2000.

Mullis, Ina V. S., M. O. Martin, E. J. Gonzalez, and S. J. Chrostowski. *TIMSS 2003 International Mathematics Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades,* Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College, 2005.

Nohara, D. *A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R),and the Programme for International Student Assessment (PISA)*. (NCES Publication No. 2001-07).Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2001.

O'Sullivan, C. Y., M. A. Lauko, W. S. Grigg, J. Qian, and J. Zhang. *The Nation's Report Card: Science, 2000.* Washington, DC: National Center for Education Statistics, 2003.

Pashley, P. J., and G. W. Phillips. *Toward World Class Standards: A Research Study Linking National and International Assessments*. Princeton, NJ: Educational Testing Service, 1993.

Wolter, K. *Introduction to Variance Estimation*. New York: Springer-Verlag, 1985.

American Institutes for Research
1000 Thomas Jefferson Street
Washington, D.C. 20007
202.000.0000
www.air.org