AMERICAN INSTITUTES FOR RESEARCH®

LESSONS LEARNED FROM U.S. INTERNATIONAL SCIENCE PERFORMANCE

PREPARED FOR:
U.S. Department of Education
Policy and Program Studies Service (PPSS)

PREPARED BY:
American Institutes for Research®
1000 Thomas Jefferson Street, NW
Washington, DC 20007-3835

December, 2007

LESSONS LEARNED FROM U.S. INTERNATIONAL SCIENCE PERFORMANCE

ALAN GINSBURG
U.S. DEPARTMENT OF EDUCATION

STEVEN LEINWAND
AMERICAN INSTITUTES FOR RESEARCH

ELIZABETH POLLOCK
AMERICAN INSTITUTES FOR RESEARCH

This paper was supported by funds from the U.S. Department of Education and The Urban Institute. The paper does not necessarily reflect the official positions of the U.S. Department of Education or The Urban Institute. The contents are the sole responsibility of the authors.

TABLE OF CONTENTS

LIST OF EXHIBITS

EXECUTIVE SUMMARY

In 2003, U.S. students' science performance was assessed by the Trends in International Mathematics and Science Study in grades 4 (TIMSS-4) and 8 (TIMSS-8) and by the Program for International Student Assessment (PISA) for students at age 15 (modal grade 10). Because all three tests are rarely given in the same year (only once every 12 years), this event offers an opportunity to compare U.S. students' science performance at different grade and age levels. When comparing the results of the U.S. with the results of the countries participating in each assessment, U.S. students score substantially above average on science assessments in grades 4 and 8 and their performance falls sharply to below average on PISA.

But by changing the composition of the countries to include only those that participated in each of the three assessments, a different picture is painted of U.S. performance. This study compares the U.S. to the 11 other industrialized countries that participated in all three science assessments. Along with the United States, the comparison countries are Australia, Belgium, Hong Kong, Hungary, Italy, Japan, Latvia, Netherlands, New Zealand, Norway, and the Russian Federation.

The international comparisons also provide a laboratory of different education systems to explore non-science achievement background variables that are theoretically related to international performance differences, as follows:

- Relative mathematics performance
- Science topic coverage
- Teacher preparation in the sciences
- Student characteristics, including gender

Note that the findings of an association from the cross-country bivariate correlations are only suggestive of a relationship and are not causal. Relationships need to be followed up with analyses particular to the factors of a particular country.

U.S. Declining Results. Based on the scores of a constant set of 12 industrialized countries, U.S. science performance exhibits successively lower rankings across assessments:

- U.S. students' average science score ranks 3rd on TIMSS-4, 5th on TIMSS-8, and 9th on PISA.
- On TIMSS-4, the score of one country statistically exceeded the U.S. score and six were below. On TIMSS-8, the scores of three countries were statistically above the U.S. score and five were below. On PISA, six countries statistically exceeded the U.S. score and none were below.

When U.S. international science performance is analyzed in relation to the performance by a comparison group consisting of a constant set of industrialized countries, U.S. performance is successively lower starting in middle school, not secondary school.

Differential U.S. Performance Across Science Content Areas. TIMSS-4 breaks out scores by life science, physical science, and earth science. TIMSS-8 further breaks out physical science into physics and chemistry and adds environmental science. On TIMSS-4, U.S. science performance is about equal across science strands. On TIMSS-8, as science knowledge deepens, U.S. students perform relatively better in biology and environmental science and relatively poorer in chemistry and physics.

Mathematics and Science Performance. Mathematics is considered the essential language to understand and communicate science understanding. A recent U.S. study found that college students who took more mathematics in high school had better grades in college science courses.

The 12 countries that participated in the three international science assessments also participated in the three international mathematics assessments given in the same year, providing an independent data source to check whether the association holds between a country's relative mathematics and science performance at the primary and secondary levels. Expectations are that students' mathematics knowledge should increase in its importance for science performance as science advances into the upper grades. Across the 12 countries, the following results were found:

- Mathematics and science rank correlations are positive, with the correlations increasing across assessments: 0.64 on TIMSS-4, 0.73 on TIMSS-8, and 0.83 on PISA, age 15.

- Mathematics and science score correlations are consistently positive and significant, although all three score correlations were consistently high (over 0.8). Score correlations, unlike rank correlations, are sensitive to outliers. Once Norway's extreme scores were eliminated, the score correlations followed the expected successive increases across the three assessments: 0.67 on TIMSS-4, 0.84 on TIMSS-8, and 0.91 on PISA, age 15.

The United States' consistently low international mathematics scores are one explanation for declining U.S. science performance in the upper grades. By contrast, Australia's and New Zealand's relative mathematics performances are higher with successive assessments and parallel their successively higher relative science performance.

Science Curricula Coverage and Standards. Students in countries with intended curricula that cover a high percentage of the topics presented on the international assessments possess an exposure advantage that works to boost scores. However, high topic exposure may also be counterproductive by giving priority to breadth of learning over depth of content understanding. Whether the benefits of broad topic exposure outweigh its disadvantages is an empirical question.

The U.S. intended science curricula, as estimated by content experts, covers the highest percentage of science topics on TIMSS-4 and the second highest percentage of topics on TIMSS-8. These high rates of topic coverage suggest a lack of depth in the U.S. science curricula that contributes to the falloff in U.S. performance across successive higher-grade assessments.

Across all countries, an increase in the percentage of TIMSS topics that a country intends to teach is negatively related to its science performance, but surprisingly no such relationship is observed between the percentage of TIMSS topics that a country actually teachers and its science performance. This may suggest the importance of having a set of science standards that represent a sound blueprint of intended topics. An analytical comparison of Japanese and Australian standards with the U.S. National Science Education Standards (science standards put forward by the National Academy of Sciences and cited by many states as guideposts for their state standards) shows how one important prototype set of U.S. science standards has less depth in its treatment of science content than the science standards in Japan and Australia.

Teachers of Science by Major. Limited available research suggests a relationship between the science preparation of teachers and the science performance of their students. However, across all 12 countries, no association was observed between the percentage of a country's grade 8 science teachers for whom science was a major area of study and its international science scores. When science teachers' majors were broken out, an association *was* found between the percentage of teachers with a science major in a particular science subject (e.g., biology) and the relative strength of a country's performance in that subject (e.g., biology).

Student Characteristics. In many countries, males are more likely to enter science occupations than females. A breakout of each country's science scores by gender fails to find significant gender differences in science scores on TIMSS-4 or PISA, but it does show boys significantly outscoring girls across most countries on TIMSS-8. Middle school is a transition period in maturity and course taking, and it may be that gender differences are prominent for this age group. Other student characteristics, such as time spent on homework and attitudes toward science, were not positively associated with a country's science performance.

Implications. Several policy considerations are presented for further consideration to improve U.S. science performance including: placing a greater emphasis on strong mathematics preparation and possibly delaying science education until grade 3 as Japan does; increasing the percentage of teachers with a science major, especially in the physical sciences; revising the National Science Education Standards to strengthen its focus and produce a deeper treatment of science topics; and making sure that girls in middle school receive appropriate encouragement. Part of the follow-up considerations should also be the launching of rigorous U.S. intervention research to evaluate how well these correlational findings can be replicated within the U.S. and to better understand the application of the international results to improve U.S. science instruction.

LESSONS LEARNED FROM U.S. INTERNATIONAL SCIENCE PERFORMANCE

Introduction

All three major international assessments in science were given during 2003, an event that occurs only once every 12 years and offers a rare opportunity to compare performance across different grade and age levels. On the grade 4 Trends in International Mathematics and Science Study (TIMSS-4), U.S. students ranked in the top quarter of all participating countries (6th of 25). On the grade 8 assessment (TIMSS-8), the U.S. ranking slightly improved to reach the top fifth of all participants (9th of 45). On the Program for International Student Assessment (PISA), which is administered to 15-year-olds, the U.S. rank against all participating countries falls precipitously below the international average (22nd of 40). On the face of it, these results suggest that U.S. science rankings exhibit a sharp falloff beginning at secondary school and that U.S. efforts to strengthen international science performance should focus on improving secondary school science.

However, the rankings of U.S. international science performance against all the countries participating on each of the three international assessments have two important limitations. First, U.S. performance is compared against the performance of countries that represent a mix of both industrialized and developing economies. A better gauge of U.S. students' science success would be to compare U.S. students' science performance to only the industrialized countries that represent U.S. economic competitors. Second, the mix of industrialized nations changes across the three international assessments. Thus, the U.S. science rankings may reflect the changing mix of countries participating in each assessment and not reflect real changes in U.S. relative performance.

The National Science Foundation discussion of the TIMSS 2003 science results (NSF, 2004) ignored the comparison group issue when it focused on U.S. 2003 science score improvement compared with the scores of all countries that participated in the prior 1995 TIMSS assessments: U.S. "fourth grade students remained fifth among 15 countries that participated in both 1995 and 2003" and "Eighth-grade students…raised their overall standing among the 21 countries that participated in TIMSS in both 1995 and 2003." However, these U.S. comparisons are based on two different country sets and include different mixes of industrialized and non-industrialized countries.

The current study corrects for weaknesses in measuring U.S. science performance by recomputing U.S. rankings against only the common set of 11 other industrialized countries participating in all three 2003 science exams. This approach of employing a common comparison group of industrial countries was previously used to assess U.S. rankings on the international mathematics assessments. It produced significantly different U.S. results compared with the full rankings against all countries participating on each assessment (Ginsburg, Cooke, Leinwand, Noell, & Pollock, 2005).

Along with examining U.S. international science performance, this study examines several country background variables that research suggests may be important in explaining students' science outcomes on the international assessments. One such variable is students' mathematics performance. A recent study of U.S. college students found that students' secondary mathematics course taking was associated with science performance in college (Sadler & Tai, 2007). The current study provides independent information on which to examine the relationship between mathematics and science performance by comparing a country's international mathematics performance with its international science performance at the elementary and secondary levels.

Along with a country's mathematics performance, three other sets of background variables are examined: science curriculum exposure, science preparation of teachers of science, and student characteristics. Because PISA rotates the emphasis of its assessments among reading, mathematics, and science, the 2003 PISA assessment collected only a limited set of science background variables. Hence, this report relies primarily on the TIMSS background data.[1]

A cautionary note is in order when interpreting the correlation between a country's science performance and its background characteristics. Large-scale international assessments cover widely different international systems and offer a natural laboratory in which to identify characteristics of science systems associated with performance differences. However, the correlations are only suggestive of an association; they are not, by themselves, evidence of causation. The bivariate correlations do not simultaneously control for the many unobserved country differences that could influence the bivariate correlations. They require further studies of the effect of these characteristics within a particular country to demonstrate their applicability.

This paper consists of six sections:

- Section I describes key design features of the TIMSS and PISA international assessments along with data limitations.

- Section II examines how the U.S. scores rank internationally across the common set of 12 countries on TIMSS at grades 4 and 8 and PISA at age 15 (modal grade 10). It also explores the stability of scores to determine whether a country's grade 4 performance is a predictor of its grade 8 performance and whether its grade 8 performance is a predictor of its performance on PISA at age 15.

- Section III examines a country's mathematics performance in relation to its science performance.

- Section IV examines a country's intended and actual coverage of topics on the TIMSS assessment in relation to its science performance. It also presents a brief qualitative comparison of science standards in the United States with those for Japan and Australia, whose performance levels differ from those of the United States.

- Section V examines whether the international science performance of the United States varies from other countries due to differences in science teacher preparation and student characteristics such as gender and attitudes toward science.

[1] A much richer set of background variables will be available from the PISA 2006 science data to be released in December 2007. However, these data represent country performance three years after the 2003 TIMSS and hence are not comparable in time.

- Section VI presents implications of the findings and suggestions for strengthening U.S. science instruction and further U.S. research to strengthen the knowledge base about the applicability of these findings to improve U.S. science instruction.

I. CHARACTERISTICS OF THE 2003 TIMSS AND PISA SCIENCE ASSESSMENTS

A valid interpretation of U.S. science performance based on TIMSS and PISA requires an understanding of the similarities and differences between these two international assessments. Exhibit 1 compares the purpose, content areas, cognitive skills, number and types of items, background questionnaires, countries participating, and assessment cycle of the two assessments.

Exhibit 1. Key Features of TIMSS Grades 4 and 8 and PISA Age 15 Science Assessments, 2003

	TIMSS	PISA
Purpose	Assess “students’ achievement in relation to different types of curricula, instructional practices, and school environments” (Mullis et al., 2004, p. 13).	Measure how well young adults are prepared to meet the challenges of today’s societies. Emphasis on the application of science knowledge and skills in real-life situations, as opposed to testing particular curricular components. Assess scientific literacy, or the capacity to use scientific knowledge, to identify questions, and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity.
Science Framework: Content Areas	• Life Science • Physical Science (only gr. 4; combined reporting of chemistry and physics topics) • Chemistry (only gr. 8) • Physics (only gr. 8) • Earth Science • Environmental Science (only gr. 8)	Limited assessment time was available for science in 2003, so it was not possible to report on major scientific fields of physics, chemistry, biological science, and earth and space science. Items were classified according to 13 major scientific themes or concepts (structure and properties of matter, atmospheric change, chemical and physical changes, energy transformations, forces and movement, form and function, human biology, physiological change, biodiversity, genetic control, ecosystems, the earth and its place in the universe, and geographical change).
Science Framework: Cognitive Skills	• Factual Knowledge • Conceptual Understanding • Reasoning and Analysis	• Describing, explaining, and predicting scientific phenomena • Understanding scientific investigation • Interpreting scientific evidence and conclusions
Assessment Items	Grade 4: 150 assessment items with 59% multiple choice and 41% constructed response Grade 8: 189 assessment items with 58% multiple choice and 42% constructed response	35 assessment items with 57% multiple choice and 43% constructed response
Background Questionnaires	Student, teacher, and principal questionnaires	Student and principal questionnaires
Country Participation	Grade 4: 25 countries with 19 industrialized (high and upper middle per capita income on 2005 data from World Bank) Grade 8: 45 countries with 32 industrialized (high and upper middle per capita income on 2005 data from World Bank)	Age 15: 40 countries with 35 industrialized (high and upper middle per capita income on 2005 data from World Bank)
Assessment Cycle	Assessment of science every 4 years	Assessment of some science every 3 years, with next in-depth science component in 2006 and next same-year assessment as TIMSS in 2015.
Sources: Mullis, Martin, Gonzalez, & Chrostowski, 2004; OECD, 2004; World Bank at: http://devdata.worldbank.org/wdi2006/contents/income.htm.		

Three significant differences should be noted in comparing results between the 2003 TIMSS and the 2003 PISA assessments. First, TIMSS is designed to measure student understanding of the science content typically taught in science classes in TIMSS participating countries through grades 4 and 8. TIMSS questions are designed to balance factual knowledge, conceptual understanding, and reasoning. By contrast, PISA questions are designed to assess students' ability to apply their science knowledge to real-life situations. As such, the PISA assessment is less about factual recall of science facts than is TIMSS and more about measuring students' ability to apply science knowledge.

Second, the 2003 PISA assessment focused most heavily on mathematics, and the PISA science assessment and background questions were more limited than those for TIMSS. However, in 2003, the PISA and TIMSS assessments were both given in the same year, something that will not happen again until 2015.

Third, the number and composition of countries taking TIMSS-4, TIMSS-8, and PISA differ. The TIMSS countries tend to be less industrialized than the PISA countries. Moreover, even between TIMSS-4 and TIMSS-8, a sharp increase in number of participating countries occurs. Hence, when comparing the U.S. performance against those of a common set of industrialized countries participating in all three assessments, it is important to achieve a constant industrialized country grouping.

II. U.S. SCIENCE PERFORMANCE ON INTERNATIONAL ASSESSMENTS

Overall U.S. Performance

To develop an appropriate common country comparison group for the United States, we identified the 11 other industrialized countries that participated in all three 2003 international assessments. An industrialized country was defined as an economy with a 2005 Gross National Income per capita that is high or upper middle income as computed by the World Bank (2006). Three countries among the 12 fall within the World Bank upper-middle-income group: Hungary, Latvia, and the Russian Federation. The remaining 9 countries are designated by the World Bank as high-income economies: Australia, Belgium, Hong Kong (China), Italy, Japan, Netherlands, New Zealand, Norway, and the United States.

Exhibit 2 displays the average country scores and country rankings for the 12 industrialized countries participating in all three of the 2003 science assessments. The TIMSS and PISA scales are comparable in that each is set at a mean of 500 and a standard deviation of 100.

A comparison of the U.S. average science score with those of the 11 other industrialized countries shows that the United States has successively lower science rankings across the three assessments. The United States ranks third on TIMSS-4 but slips to fifth on TIMSS-8. By age 15 (modal grade 10), the science scores of U.S. students have fallen to eighth. This successive decline in U.S. rankings on the three assessments against a common set of industrialized countries differs from the U.S. rankings against all participating countries, which show a precipitous reduction only on PISA after some improvement in the U.S. ranking between TIMSS-4 and TIMSS-8.

Exhibit 2. Scores and Rankings of 12 Industrialized Countries Participating on the 2003 International Science Assessments[1]: TIMSS Grades 4 and 8, and PISA Age 15

Country	TIMSS Grade 4		TIMSS Grade 8		PISA Age 15	
	Score	Rank[2]	Score	Rank[2]	Score	Rank[2]
AUS	521[a]	8	527	5	525[a]	3
BEL	518[a]	10	516[a]	8	509[a]	6
HKG	542	2	556[a]	1	539[a]	2
HUN	530	5	543[a]	3	503	7
ITL	516[a]	11	491[a]	12	486	11
JPN	543[a]	1	552[a]	2	548[a]	1
LAT	532	4	512[a]	10	489	9
NLD	525[a]	7	536	4	524[a]	4
NZL	520[a]	9	520	7	521[a]	5
NOR	466[a]	12	494[a]	11	484	12
RUS	526	6	514[a]	9	489	9
USA	536	3	527	5	491	8
	Countries statistically above U.S. = 1 Countries statistically below U.S. = 6		Countries statistically above U.S. = 3 Countries statistically below U.S. = 5		Countries statistically above U.S. = 6 Countries statistically below U.S. = 0	

	Cor. Among Country Science Scores	Cor. Among Country Science Ranks
TIMSS-4 and TIMSS-8	0.71**	0.71**
TIMSS-8 and PISA-Age 15	0.80**	0.86**
TIMSS-4 and PISA-Age 15	0.49*	0.52*

[a] Indicates country scored statistically significant above or below the United States at the .05 level.*Indicates significant at .1 level; ** indicates significant at .05 level.

[1]Tunisia also participated in all three international results, but it is not an industrialized country and was omitted from our study. England was also omitted because it did not satisfy guidelines for participation.[2]Country rankings are from highest score (equals 1) to lowest score (equals 12).Source: Mullis, Martin, Gonzalez, & Chrostowski, 2004; OECD, 2004.

A steady reduction in U.S. relative performance is also observed when the comparison is limited to science scores statistically different from the U.S. science score. For grade 4, only Japan's science score is significantly above the U.S. score, and the science scores of six countries are significantly below the U.S. science score. For grade 8, three countries, Japan, Hong Kong, and Hungary, score significantly above the United States and five countries score below the United States. For students at age 15, six countries' science scores significantly exceed the U.S. score, and no country's score is statistically significant below the U.S. score.

Along with the constancy of U.S. science performance across the three assessments, the constancy of performance can be measured for all 12 countries. The correlation coefficient indicates the strength of association on scores or ranks between TIMSS-4 and TIMSS-8, between TIMSS-8 and PISA, and between TIMSS-4 and PISA. A correlation of near 1 is very strong and a correlation near zero is very weak. Both score and rank correlations between TIMSS-4 and TIMSS-8 are a moderate 0.7, which means that about half the variance in TIMSS-8 scores or ranks is associated with the variation in TIMSS-4. The correlations between TIMSS-8 and PISA are a somewhat stronger 0.8, meaning that about two-thirds of the variance in PISA scores or ranks across countries is associated with the variation in TIMSS-8. Overall, the score and rank correlations between TIMSS-4 and PISA scores are about 0.5. Thus, considerable variation may occur between a country's initial science performance and its later science performance.

The increase in rank shown by Australia's and New Zealand's science performance, two westernized countries with systems somewhat similar to that of the United States, contrasts with the decrease in rank for U.S. performance across the three assessments. Australia ranks 8th on TIMSS-4, moves up to 5th on TIMSS-8, and achieves 3rd on PISA among the 12 countries. New Zealand's ranking moves up from 9th to 7th to 5th, respectively. We shall further explore the possible reasons for these countries' improvements as contrasted with the decline in U.S. ranking across the three assessments.

Relative Country Performance Across Science Content Areas

Exhibit 3 displays a country's relative strengths and weaknesses by science content area.[2] Grade 4 scores are broken out for life science, physical science, and earth science. Grade 8 scores further break out the physical sciences into chemistry and physics and include an additional category of environmental science. A country's relative strength or weakness in a content area is measured as the difference between its average score for a science content area and its average score across all science content areas. A positive difference indicates that a country performs better than its average science score. A negative difference indicates that it performs below its country average science score.

On TIMSS-4, U.S. performance across the content areas is relatively uniform. No content area differs from the U.S. performance average by more than 3 points, and none of these differences is statistically significant from the average U.S. score. On TIMSS-8, large differences emerge between the U.S. content area performances and the U.S. average. The U.S. strengths are in life science (plus 8 points) and environmental science (plus 7 points), and the U.S. weakness is in chemistry (minus 13 points). Performance on physics is also much weaker than the U.S. average (minus 11 points), although the negative difference is not statistically significant. Across all countries, the correlations between relative performance on TIMSS-4 and TIMSS-8 within a content area are not statistically significant except for the physical sciences. Hence, content-area performance on TIMSS-4 is not a strong predictor of content area performance on TIMSS-8.

The differential U.S. performance by science content area on TIMSS-8 offers one source for the decline in U.S. score performance between TIMSS-4 and TIMSS-8. The TIMSS-8 assessments include relatively fewer life science items compared with TIMSS-4. Only 29 percent of the TIMSS-8 items are in life science compared with 43 percent of the TIMSS-4 items. But U.S. students have a relative performance strength in the life sciences, so that its reduced emphasis on TIMSS-8 produces a score reduction compared to a TIMSS-8 with life sciences weighted as in TIMSS-4.

[2] The abbreviated science part of the PISA 2003 assessment asked too few questions to break out the distribution of country scores by science content areas.

Exhibit 3. Difference in the Average Scale Scores for Science Content Area from the Average Scale Scores Across All Content Areas on TIMSS 4 and TIMSS 8 for the 13 Countries Participating in All Three International Tests

	Life Science		Physical Science			Earth Science		Environ-mental Science
				Chemistry	Physics			
Country	Gr. 4	Gr. 8	Gr. 4	Gr. 8	Gr. 8	Gr. 4	Gr. 8	Gr. 8
AUS	3	7	–2	–19**	–4	–2	6	11**
BEL	6*	11**	–11*	–12**	–1	4**	–7	8**
HKG	–5	1	8*	–8	5	–4	–1	5
HUN	7*	–3	–3	21**	–3	–3	–2	–11
ITL	4	5	–5	–6	–23**	2	20**	4
JPN	–11*	3	16*	6*	18**	–6	–16**	–9**
LAT	–1	–1	0	2	0	2	2	–4
NLD	29*	4	–13*	–18**	6	–15**	2	7
NZL	1	5	–3	–17**	–3	3	7	7
NOR	10*	0	–14*	–11	–	3	21**	0
RUS	–1	2	0	15**	–1	0	6	–21**
USA	3	11**	–3	–13**	–11	1	6	7**
	Correlation among country life science mean scores for TIMSS-4 and TIMSS-8 = .07		Correlation among country physics mean scores for TIMSS-4 and TIMSS-8 = .52** Correlation among country chemistry mean scores for TIMSS-4 and TIMSS-8 = .40*			Correlation among country earth science mean scores for TIMSS-4 and TIMSS-8 = .37**		

[a] Indicates country scored statistically significant above or below the United States at the .05 level.

*Indicates Pearson correlation for scores or Spearman correlation for ranks is significant at .1 level; ** indicates significant at .05 level.

Source: Mullis, Martin, Gonzalez, & Chrostowski, 2004; OECD, 2004.

III. MATHEMATICS AND SCIENCE RANKINGS

Four centuries ago (1623), Galileo wrote about how the language of mathematics is the language essential to understanding science (Machamer, 1998).

> *Philosophy is written in this grand book, the universe which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth."*
>
> —Galileo Galilee in *Assayer*

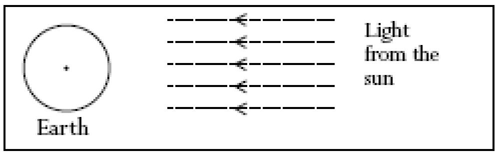
A recent study of college undergraduates found that exposure to more high school mathematics was associated with achieving better grades in college science courses, controlling for other factors such as high school grade point average and students' socio-economic backgrounds (Sadler and Tai, 2007). A multiple regression found that the contribution of high school mathematics courses to college grades was as large as the contribution of prior high school sciences courses in the same science topic to college grades in that topic. Qualifications to these findings included that prior high school coursework was based on recall, the results were quasi-experimental, and the sample was limited to students attending liberal arts colleges.

The three international assessments offer another data source from which to test the importance of learning mathematics for science performance, this time at elementary and secondary levels. The 12 countries participating in the three international science assessments also participated in the three international mathematics assessments administered in the same year. In the elementary grades, science tends to involve factual recall, so students' mathematics performance would be expected to increase in importance as a factor determining science performance in the upper grades, when students have to employ more mathematics thinking to explain science concepts.

A sample question on the 2003 physics portion of the PISA exam illustrates the application of mathematical thinking at the secondary level.

Today, as the Northern Hemisphere celebrates its longest day, Australians will experience their shortest. In Melbourne, Australia, the sun will rise at 7:36 am and set at 5:08 pm, giving nine hours and 32 minutes of daylight. Compare today to the year's longest day in the Southern Hemisphere, expected on 22 December, when the sun will rise at 5:55 am and set at 8:42 pm, giving 14 hours and 47 minutes of daylight. The President of the Astronomical Society, Mr. Perry Vlahos, said the existence of changing seasons in the Northern and Southern Hemispheres was linked to the Earth's 23-degree tilt.*

In the Figure light rays from the sun are shown shining on the earth. Suppose it is the shortest day in Melbourne. Show the earth's axis, the Northern Hemisphere, the Southern Hemisphere and the Equator on the figure. Label all parts of your answer.



Full score: Answers which include a diagram with the Equator tilted towards the sun at an angle between 10° and 45° and the earth's axis tilted towards the sun within the range 10° and 45° from vertical, and the Northern and or Southern Hemispheres correctly labeled (or one only labeled, the other implied).

This is an open-response item that requires students to create a conceptual model in the form of a diagram showing the relationship between the rotation of the earth on its tilted axis and its orientation to the sun on the shortest day for a city in the southern hemisphere. In addition they had to include in this diagram the position of the equator at a 90-degree angle to the tilted axis. Full credit is obtained if the students correctly place and label all three significant elements – the hemispheres, the tilted axis and the equator. Partial credit is given for a diagram with two of the three elements correctly placed and labeled.

The sample PISA problem does not require students to understand advanced high school mathematics, such as trigonometry or calculus, but does require students to understand how to apply fundamental mathematical ideas about the geometry of angles.

Exhibit 4 displays each country's science and mathematics rankings for TIMSS-4, TIMSS-8, and PISA. The United States ranks 3rd in science on TIMSS-4, substantially exceeding its rank of 8th in mathematics. By age 15 on PISA, U.S. students' rankings on science and mathematics are nearly identical—8th on science and 9th on mathematics. This pattern suggests that mathematics becomes an increasingly influential factor in determining science performance in the upper grades, as would be expected.

Exhibit 4. Rankings of 12 Industrialized Countries Participating on the 2003 International Science and Mathematics Assessments[1]: TIMSS Grades 4 and 8, and PISA Age 15

	TIMSS Grade 4		TIMSS Grade 8		PISA Age 15	
Country	Science Rank	Mathematics Rank	Science Rank	Mathematics Rank	Science Rank	Mathematics Rank
AUS	8	10	5	8	3	5
BEL	10	3	8	3	6	4
HKG	2	1	1	1	2	1
HUN	5	7	3	5	7	8
ITL	11	9	12	11	11	12
JPN	1	2	2	2	1	3
LAT	4	5	10	6	9	9
NLD	7	4	4	4	4	2
NZL	9	11	7	10	5	6
NOR	12	12	11	12	12	7
RUS	6	6	9	6	9	11
USA	3	8	5	9	8	9

	Cor. Among Country Science and Math Scores	Cor. Among Country Science and Math Ranks
TIMSS-4	.83**	.64**
TIMSS-8	.87**	.73**
PISA-Age 15	.89*	.83**

*Indicates Pearson correlation for scores or Spearman correlation for ranks is significant at .1 level; ** indicates significant at .05 level.

[1]Tunisia also participated in all three international results, but it is not an industrialized country and was omitted from our study. England was also omitted because it did not satisfy guidelines for participation.

[2]Country rankings are from highest score (equals 1) to lowest score (equals 12).

Source: Mullis, Martin, Gonzalez, & Chrostowski, 2004; OECD, 2004.

Across all countries, an increasingly higher correlation in the upper grades would be expected given that mathematics explanations are more important for upper-grade science. The association between a country's mathematics and science rankings becomes higher over the three assessments on the rank correlation. At grade 4, the rank correlation between mathematics and science scores is statistically significant, but a modest 0.64. At grade 8, the rank correlation is a higher 0.73. At age 15 on PISA, the rank correlation is still higher at 0.83, with mathematics rankings explaining about two-thirds of the variation in science rankings.

The mean-score correlation increases between TIMSS-4 and PISA, but the increase is very small. Unlike the rank correlations, the score correlations are sensitive to outliers. Norway has the lowest science score among the 12 countries, one that is 50 points below the next lowest score, and Norway's mathematics score is 40 points below the next lowest mathematics score. Eliminating Norway's extreme scores from the correlation computations yields the following mean-score correlations between mathematics and science:

- 0.67 on TIMSS-4
- 0.84 on TIMSS-8

- 0.91 on PISA at age 15

Thus, once the extreme Norway outlier is removed, the association between mathematics and science scores follows the expected increase in correlational strength at higher grade and age levels.

Moreover, one possible hypothesis that might explain the improvements in Australia's and New Zealand's science scores is their improvement in their mathematics scores. Between TIMSS-4 and PISA, Australia improves its international mathematics rank by five positions and its science rank also goes up by five positions. New Zealand's mathematics rank improves between TIMSS-4 and PISA by five positions and its science rank by four positions.

Overall, the international correlational evidence supports the idea that mathematics performance is important for science performance especially in the upper grades. It is important to validate the international correlational findings for the U.S. through rigorous impact evaluations. Such studies would evaluate interventions designed to improve students' science performance by strengthening their mathematics skills and the use of mathematics in the science curricula.

IV. SCIENCE CURRICULA: COVERAGE AND STANDARDS

Exposure to Science Topics

Students won't learn what they are not taught. Using this logic, policymakers might expect that a country's science ranking would be positively related to its students' exposure to science topics. That is, countries that teach a higher percentage of the TIMSS intended content should have an exposure advantage that produces higher scores.

But cognitive research suggests a different scenario. Students are better able to understand and apply what they are taught when instruction teaches concepts in depth and with rich applications that develop conceptual understanding (AAAS, 2004; Bransford et al., 2000; Brunner, 1960). This line of reasoning suggests that countries with high topic exposure will have a negative association with learning and lower scores.

Prior studies from the 1995 TIMSS assessment concluded that, in fact, topic depth was lacking in the U.S. curricula. The U.S. science and mathematics curricula were characterized as "a mile wide and an inch deep" (Schmidt and McKnight, 1998) and that the teaching of too many TIMSS topics was a source of low U.S. scores.[3] The current analysis examines whether this association still holds for the 2003 TIMSS.

Exhibit 5 displays the percentage of science topics covered on the TIMSS-4 and TIMSS-8 assessments that each country intends to teach and actually does teach. Experts from each country estimated the intended content-exposure percentages for their country, and teachers filling out the survey forms estimated the percentages of topics they actually teach.

[3] A report by the National Research Council of the National Academy of Sciences (2002) found that even at the high school level, U.S. Advanced Placement courses had "excessive breadth of coverage (especially in 1-year science programs) and insufficient emphasis on key concepts in final assessments."

Exhibit 5. Percentage of TIMSS Topics Intended to Be Taught and Actually Taught Through Grade 4 and Grade 8

	TIMSS Topics Intended to be Taught (Pct. of Total)				TIMSS Topics Actually Taught (Pct. of Total)			
	Grade 4 (32 Topics)		Grade 8 (44 Topics)		Grade 4 (32 Topics)		Grade 8 (44 Topics)	
Country	Pct	Rank	Pct	Rank	Pct	Rank	Pct	Rank
AUS	63	5	55	11	55	6	52	7
BEL	31	12	23	12	43	9	48	9
HKG	53	8	64	9	62	4	53	6
HUN	50	9	91	3	71	1	84	1
ITL	69	3	98	1	65	3	77	3
JPN	50	9	73	7	37	10	52	7
LAT	41	11	82	5	—		64	4
NLD	69	3	73	7	47	8	58	5
NZL	59	6	57	10	62	4	45	10
NOR	78	1	80	6	55	7	45	10
RUS	56	7	86	4	—			
USA	78	1	95	2	69	2	79	2

	Cor. Among Country Science Scores and Pct. of TIMSS Topics Intended to Be Taught	Cor. Among Country Science Ranks and Pct. of TIMSS Topics Intended to be Taught		Cor. Among Country Science Scores and Pct. of TIMSS Topics Actually Taught	Cor. Among Country Science Ranks and Pct. of TIMSS Topics Actually Taught
TIMSS-4	–.41*	–.31	TIMSS-4	.02	.04
TIMSS-8	–.16	.02	TIMSS-8	.01	.18

*Indicates Pearson correlation for scores or Spearman correlation for ranks is significant at .1 level; ** indicates significant at .05 level.

Source: Source: Mullis, Martin, Gonzalez, & Chrostowski, 2004.

For 2003, the United States continues the same pattern of a high level of intended topic exposure that was found in the 1995 TIMSS. On TIMSS-4, 78 percent of the science topics are intended to be taught, a coverage rate that ties the United States for the highest coverage with low-performing Norway. On TIMSS-8, the intended U.S. curricula covers 95 percent of all topics, second only to low-performing Italy. The high U.S. rates of intended exposure to TIMSS-4 topics carries over to actual coverage as reported by U.S. teachers. On TIMSS-4, the actual U.S. coverage rate ranks second highest at 69 percent, and it also ranks second highest on TIMSS-8 at 79 percent.

If high U.S. rates of topic exposure produce superficial science learning, then a question to ask is why U.S. students rank high on TIMSS-4 and then decline on TIMSS-8. One possible explanation is that important content differences occur between grade 4 and grade 8 in the composition of the TIMSS science assessment items. Along with the decreased emphasis on life science questions, as noted above, the TIMSS-8 assessment increases its emphasis on items requiring science reasoning (23 percent of TIMSS grade 4 items assess science reasoning and analysis compared with 31 percent of the grade 8 items). Accordingly, the lower U.S. rankings on TIMSS-8 may in part be attributed to the increase in the percentage of the more difficult reasoning and analysis questions. This suggests that early science instruction may not build a strong foundation for future analytic science learning.

Another explanation is that U.S. children are exposed to more informal science learning out of school, including through television and museums. Hard evidence is lacking on informal science learning, but the National Academy of Sciences (in progress) is currently synthesizing research on informal science learning.

The rank order correlation coefficient in Exhibit 5 summarizes the relationship between the percentage of TIMSS topics a country intended to teach or actually taught and its science scores. With respect to the intended curricula, although neither the TIMSS-4 nor the TIMSS-8 correlations were statistically significant, the correlations were negative, indicating that greater rates of intended topic coverage were associated with lower, not higher, country science scores.

However, a country's percentage of TIMSS topics actually taught shows a low and non-significant correlation with science scores. Although high-performing Japan has below average rates of coverage, high-scoring Hungary ranks number one in topic coverage. The two countries with the lowest actual rates of coverage, New Zealand and Norway, score relatively low on the TIMSS-4 and TIMSS-8 science assessments.

These findings suggest a more complicated relationship between science topic coverage and science performance than did the prior TIMSS research. Intended topic coverage is not the same as actual topic coverage. Most important, actual topic coverage does not directly translate to assessment performance. High-performing Hong Kong and Japan tend to fall around the middle rankings of the topic coverage rankings, indicating a parabolic relationship that requires a country to achieve a balance between covering too many or too few topics. The results further suggest that the nature of the science curricula mediates between topic coverage and scores. To further understand this relationship, the following section briefly compares U.S. science standards with those of high-performing Japan and improving Australia.

A Comparison of Science Standards From Japan, Australia, and the United States

To better understand the underlying connection between U.S. topic coverage and performance, this section presents a qualitative comparison of the structure and content of science standards in the United States, Japan, and Australia. Japan consistently outperforms the United States, and Australia improves to overtake the United States on PISA. Because a thorough review of standards is a large undertaking beyond the scope of this study, this overview is only an initial analysis to get beneath the numbers to understand the nature of the intended scope and depth of science content taught in three countries, as defined by their science standards.

Identifying the standards for Japan is straightforward because Japan has centralized standards. Australia does not have official national standards. The standards from Victoria were chosen because Victoria is one of Australia's largest states and the standards are readily accessible over the Web. The United States has no official national science standards, but the National Science Education Standards (NSES), which the National Academy of Sciences (1996) developed, are frequently cited by many U.S. states (e.g., *Science Content Standards for*

California Public Schools, Kindergarten Through Grade Twelve, 1998) as a model for developing their science standards.

The standards are evaluated in terms of their overall structure, including starting grade, grade ranges, organization, and content description. We also examine the treatment of electricity in elementary and middle school to better understand to what degree the curricula of each country teach science topics in depth. Electricity was chosen because it is a common topic in the physical sciences, where U.S. performance is relatively weak.

The science standards of the United States, Australia, and Japan are described in greater detail in the Appendix from which we draw the following comparisons:

- Starting grades. The U.S.-NSES science standards start with kindergarten and cover all science strands. Japan does not start teaching science until grade 3, electing to use the introductory years to concentrate instruction on the Japanese language and mathematics. Australia starts teaching science earlier, as the United States does, but its science curriculum is explicitly very introductory and is not intended to cover each strand.
  Grade specificity. The U.S.-NSES standards are organized around grade bands K–4, 5–8, and 9–12, each encompassing a broad range of grades. The Japanese develop grade-specific standards in grades 3–6, specifying an introductory sequence of science topics. The Japanese shift to grade bands for middle and high school. The Australians organize content by six levels of student achievement, with each level covering approximately two grades, rather narrow compared with the U.S. four-grade range.

- Focus on science content. The U.S.-NSES cover six sets of standards: science teaching, science professional development, science assessment, science content, science education program, and science education system standards, with the science content standards in the middle of the list. The Japanese science standards focus only on science content. The Australian standards are content driven, starting with the intended science curriculum, followed by expected student learning outcomes and ending with assessment indicators to know whether students have learned the curriculum.

- Treatment of electricity. A close look at one science topic, electricity, shows that the U.S.-NSES treatment of electricity K–8 is superficial and fails to explicitly develop an understanding of the basics of electrical circuits and flows. Both the Japanese and the Australian standards systematically build up the properties and uses of electricity, its relationship with magnetism, and the essential characteristics of simple electrical circuits.

This brief review suggests that the early introduction of science topics in the U.S.-NSES standards may give U.S. students a starting edge but that the advantage is temporary. The Japanese elect to emphasize mathematics in the early years, which builds a strong mathematical foundation for long-term success. The generality of the NSES standards encourage broad but superficial treatment of content, as seen in the shallow development of U.S. electricity content compared with its treatment in the Japanese and Australian standards. The general nature of the NSES standards fails to provide much guidance to U.S. states and is a possible cause of why U.S.

intended content covers so large a number of science topics compared with that of other countries. A particular failing of the U.S. standards is that they do not build up a rich treatment of science content over the grades.

The weaker focus on developing rich science content in U.S.-NSES science education standards compared with the standards of Japan and Australia is consistent with the TIMSS video findings comparing science instruction among the three countries. Science "lessons in Australia and Japan used activities and evidence to build a coherent content storyline in which ideas and activities were carefully sequenced and explicitly connected.… In the United States, science content was often secondary to activities." (Roth, 2007) The lack of focus of U.S. standards on in-depth understanding of scientific principles compared with the Australian and Japanese standards is one plausible explanation for lower U.S. science performance in the upper grades.

V. TEACHERS AND STUDENTS

Teachers of Science

A cross-country examination of science teacher preparation and its relationship to science scores is supported by research findings that teacher content and pedagogical preparation are associated with improved student learning (Cohen & Hill, 1995; Darling-Hammond, 2000; Ferguson & Womack, 1993, Wilson, Floden, & Ferrini-Mundy, 2001). However, these studies do not specifically cover science teacher preparation for which effectiveness studies are limited. Monk's (1994) longitudinal analyses of nearly 3,000 secondary school students did find that students' science achievement was positively related to the number of science courses a teacher took in the subject area being taught and to the teacher's course work in science pedagogy.

Given that the grade 4 teachers surveyed by TIMSS were probably generalists and not subject-matter specialists and that PISA did not collect information on teacher content background, the present analysis is limited to the TIMSS grade 8 data.

Exhibit 6 displays each country's overall percentage of TIMSS-8 students taught by teachers who have science as a major area of postsecondary study. Only 58 percent of U.S. grade 8 teachers of science had science as a major area of preparation, placing the United States 11th among the 12 countries in the percentage of grade 8 teachers of science with content preparation in science. Only low-scoring Norway at 52 percent had a lower concentration of teachers with a science background, and low-scoring Italy was just above the U.S. concentration.

Nevertheless, across all 12 countries, the association between the percentage of a country's teachers with science preparation and a country's science mean score rank is only 0.1. An example of the lack of association is Russia, a country with the highest percentage of its teachers with a science major, yet Russia performs in the bottom half of countries on TIMSS-8.

One possible explanation is that very low percentages of teachers with a science background clearly will not produce high performance and higher percentages of teachers with a science background are necessary but not sufficient to produce higher scores. For example, the research studies noted the importance of pedagogical preparation as well as content preparation.

A second possibility is that science teacher preparation exerts a greater influence over science scores only in the area of a teacher's science background. Thus, a further refinement of the analysis is to break out a country's distribution of teachers with science preparation by the percentages with a biology, chemistry, or physics major area of study (Exhibit 6).[4] Among U.S. teachers with science preparation, biology dominates, with 79 percent of the teachers with science preparation having biology as a major concentration. The United States ranks 3rd highest among the 12 countries in the proportion of its prepared teachers who have biology as a major concentration. By contrast, U.S. teachers have relatively less preparation in physics and chemistry, where the United States ranks 9th and 8th, respectively, out of 12.

[4] Although TIMSS also collects information about teacher preparation in earth science, several countries in our sample did not provide this information, so comparable rankings could not be computed.

Exhibit 6. Percentages of Students Taught by Teachers With Science as a Major Area of Study

		Biology, Physics, Chemistry, or Earth Science Majors[a]		Biology Major		Chemistry Major		Physics Major	
Country	TIMSS Grade 8 Achieve-ment Rank	As a Pct. of All Teachers	Rank	As a Pct of Teachers With a Science Major	Rank	As a Pct of Teachers With a Science Major	Rank	As a Pct of Teachers With a Science Major	Rank
AUS	5	80	6	75	4	65	1	30	7
BEL	8	72	8	88	1	60	3	51	1
HKG	1	71	9	52	8	52	5	48	2
HUN	3	84	5	46	10	31	10	24	9
ITL	12	65	10	83	2	05	12	09	12
JPN	2	89	4	39	11	47	6	37	4
LAT	10	97	2	60	7	64	2	45	3
NLD	4	74	7	39	11	22	11	22	11
NZL	7	90	3	66	5	59	4	34	5
NOR	11	52	12	62	6	44	7	31	6
RUS	9	98	1	49	9	43	8	27	8
USA	5	58	11	79	3	43	8	24	9
		Correlation among country pct. of science teachers with a science major and mean science scores for TIMSS-8 = .26 Correlation among country rank on pct. of science teachers with a science major and its rank on mean science scores for TIMSS-8 = .09		Correlation among country pct. of biology teachers with a biology major and relative biology score for TIMSS-8 = .69** Correlation among country rank on pct. of biology teachers with a biology major and its rank on relative biology score for TIMSS-8 =.65**		Correlation among country pct. of chemistry teachers with a chemistry major and relative chemistry score for TIMSS-8 = –.18 Correlation among country rank on pct. of chemistry teachers with a chemistry major and its rank on relative chemistry scores for TIMSS-8 =.–.36		Correlation among country pct. of physics teachers with a physics major and relative physics score for TIMSS-8 = .56** Correlation among country rank on pct. of physics teachers with a physics major and its rank on relative physics scores for TIMSS-8 =.–.36	

Note: Teachers may indicate multiple science areas as major areas of study.

[a]Not all countries reported their teacher preparation by earth science, so the rankings would not be comparable and this breakout is omitted.

*Indicates Pearson correlation for scores or Spearman correlation for ranks is significant at .1 level; ** indicates significant at .05 level.

Source: Mullis, Martin, Gonzalez, and Chrostowski, 2004.

Next, the concentration of U.S. teachers prepared in biology, physics, and chemistry is compared with U.S. relative performance in these three topics. U.S. performance is above average in biology, where the concentration of U.S. teachers is greatest. U.S. performance is below average in physics and chemistry, where U.S. teacher preparation ranks low. This suggests that increasing the percentages of U.S. science teachers who have a major area in physics or chemistry might improve U.S. performance in these areas. However, increasing the proportion of teachers with a physics or a chemistry major does not necessarily require reducing the number of teachers with a biology specialty, inasmuch as only 58 percent of U.S. teachers have any science major.

Across all countries, the correlation coefficient shows the association between a country's ranking on its concentration of teachers with preparation in a science subject and its ranking on that subject on the TIMSS-8 science assessment. As expected, the percentages of a country's teachers with a science major in biology or physics are positively associated with that country's relative performance advantage in these areas. The score correlation between a country's percentage of teachers with a science major who have a concentration in biology and a country's biology score advantage is a statistically significant 0.69; for physics, the similar correlation is a significant 0.56 (Exhibit 6). The correlation for chemistry is a negative 0.18, but it is not statistically significant from zero. The reason for insignificance for chemistry is not clear.

In summary, our results do not indicate a simple association between teacher content preparation and country performance. Countries with low rates of teachers with science majors tend to score poorly, or in the case of the United States, exhibit declining performance, but countries with high rates of science majors as teachers are not necessarily successful. The results suggest paying attention to giving teachers science content preparation in the fields they are teaching. The United States is especially in need of more teachers with a specialty in the physical sciences, although teachers can have multiple science specialties and biology specialization is also associated with improved student performance in biology.

Student Characteristics

This section examines whether student gender and attitudes toward science are associated with science assessment achievement scores. Science has traditionally been a profession that males are more likely to enter than females (American Association of University Women, 2000). Exhibit 7 examines the differences between the science scores of girls and boys for each country on each of the three assessments.

The size of the gender differences changes across the three international assessments. On TIMSS-4, girls' and boys' scores are about equal. None of the countries experiences score differences greater than 10 points (one-tenth of a standard deviation), and in half the cases, girls outscore boys. On TIMSS-8, gender differences increase, with boys having a clear score advantage over girls. Boys outscore girls in all countries, and two-thirds of the differences exceed one-tenth of a standard deviation. On PISA, gender differences in average science scores narrow again. Boys have no score advantage in five countries, and Japanese girls outperform boys from all countries except Japan. New Zealand and Russia are the only countries in which the boys' score advantage is statistically significant. The source of girls' middle school disadvantage compared with the results from other grades warrants further attention because middle school is an important period in shaping girls' subject area interests.

This study also examined several other student characteristics that theoretically could be related to a country's science performance, but no such relationship was observed for the United States or over all 12 countries on TIMSS-4 or TIMSS-8 (PISA data were unavailable). With respect to the average students' time spent on homework, students in high-scoring Japan and Hong Kong had the highest proportions of students with low rates of time spent on homework (no more than 30 minutes of science homework no more than twice a week). The United States, along with low-scoring Italy and Norway, had the lowest proportions of students with low rates of time spent on homework. However, the data do not capture Japan's and Hong Kong's large-scale private after-school tutoring programs (e.g., Jukus), which may account for the TIMSS low-homework-time findings.

Exhibit 7. Average Scale Scores of Girls and Boys and Average Score Differences

Coun-try	TIMSS-4				TIMSS-8				PISA			
	Girls	Boys	G–B	Rank	Girls	Boys	G–B	Rank	Girls	Boys	G–B	Rank
AUS	522	519	3	3	517	537	–20*	10	525	525	0	4
BEL	518	519	–1	7	505	528	–23*	11	509	509	0	4
HKG	544	541	3	3	552	561	–9*	3	504	503	1	3
HUN	527	533	–6	11	530	556	–26*	12	541	538	3	2
ITL	514	517	–3	8	486	496	–10*	5	484	490	–6	10
JPN	542	545	–3	8	548	557	–9*	3	546	550	–4	7
LAT	534	529	5	2	509	516	–7*	1	491	487	4	1
NLD	521	529	–8*	12	528	543	–15*	8	522	527	–5	8
NZL	523	517	6	1	515	525	–10	5	513	529	–16*	12
NOR	467	466	1	5	490	498	–8*	2	483	485	–2	6
RUS	527	526	1	5	508	519	–11*	7	485	494	–9*	11
USA	533	538	–5*	10	519	536	–17*	9	489	494	–5	8

*Statistically significant at .05 level (according to TIMSS 2003 International Science Report)
(Positive G-B differences indicate girls scored higher.)

Positive student attitudes toward science were not associated with positive scores. Japan and Hong Kong had the lowest percentages of students who valued science or enjoyed science, despite high average scores. The United States had among the top percentages of students with favorable attitudes toward science, although U.S. science scores were significantly below those of Japan and Hong Kong. The negative attitudes of students in Japan and Hong Kong may reflect the pressures they experience to do well in science that are not present to the same degree in the United States.

VI. CONCLUSIONS AND REFORM DIRECTIONS

In relation to the scores of students in the other industrialized countries participating in all three 2003 international science assessments, the scores of U.S. students rank high on TIMSS-4 and rank successively lower on TIMSS-8 and PISA. This pattern of successively lower performance across assessments differs from the precipitous pattern of secondary-school falloff in the U.S. rankings that occurs when the United States is compared with all other test-taking countries. These results suggest that teaching science early affords no long-term advantage, probably because of the weak nature of science teaching in the early elementary grades.

The international comparisons provide correlational evidence to suggest the causes of the successively lower performance of U.S. students on TIMSS-8 and PISA. Six directions for reform follow from the empirical analyses to strengthen U.S. science performance and are presented below for *further consideration and validation*:

1. Strengthen U.S. mathematics performance as a strategy to strengthen U.S. science performance. The international evidence provides additional independent evidence of an association between a student's mathematics performance and their science achievement. Our analyses observed a strong correlation between a country's mathematics performance and its science performance. As theoretically expected, the strength of the correlation increased at higher grades when science requires greater mathematical thinking. These findings about the importance of mathematics suggest that U.S. students' poor mathematics performance places U.S. students at a serious disadvantage in science.

A comprehensive examination of strategies for strengthening U.S. mathematics performance is outside the scope of this paper,[5] but one strategy might be to start teaching science later and use the additional time to emphasize mathematics. As our examination of the Japanese science standards showed, the Japanese do not teach science until grade 3. Instead, they prefer to emphasize mathematics and language. The fact that the Japanese so readily catch up in science even when starting later suggests that U.S. students might benefit from a greater earlier emphasis on mathematics in grades 1 and 2, with science starting in grade 3.

2. Strengthen instruction in the physical sciences, the weakest area of U.S. science performance across the content areas. U.S. students' science performance is weaker in the physical sciences (physics and chemistry) than in other content areas. The physical sciences are particularly dependent on mathematical thinking, and improving mathematics as suggested above could also support students' learning of the physical sciences. The United States also had one of the lowest proportions of teachers with a science background whose preparation was in the physical sciences. Recruiting and financial incentives to draw more teachers into science with a preparation in the physical sciences could help.

3. Increase the overall percentage of science teachers with science as a major area of postsecondary preparation. The United States has one of the lowest percentages of grade 8 teachers of science who have science of any type as a major area of postsecondary preparation.

---

[5] For suggestions to improve U.S. students' mathematics performance, see *Adding It Up: Helping Children Learn Mathematics* (2001). National Academy Press.

Although the association between teachers' having a major in science and students' performance is not strong overall, the association between the proportion of a country's teachers who major in a particular field of science and students' relative performance in that field is statistically significant. The No Child Left Behind legislation requires that teachers be certified in the field they are teaching. Districts should place a high priority on recruiting teachers with a science background in the field they teach. Middle school certification of science teachers should consider adequate mastery of the content of individual science strands.

4. The National Academy of Sciences should consider revising the National Science Education Standards (NSES) to strengthen their content focus and specificity. The NSES are the defacto U.S. standards that many states model in specifying their own standards. The NSES were innovative for their time, when formal state science standards and aligned assessments were not widespread and before NCLB required all states to develop science standards and align their assessments. The challenge now is to develop models of good standards that the states can follow. The pilot comparisons suggest that compared with the standards of high-performing Japan and Australia, the NSES are far too general to guide the development of a coherent program of science. A revision of the NSES to improve their content focus and specificity and to reduce grade-band size would make them consistent with our international comparisons.

5. Reinforce girls' exposure to science in the middle grades. On TIMSS-4 and PISA, the differences between the science performance of girls and boys are not large, but for some reason, girls' performance consistently falls below that of boys on TIMSS-8. The reasons for the science gap for middle school girls are not clear, but the phenomenon is widespread across the industrialized countries. Middle school science programs might want to place some priority on ensuring that all students learn the science they need to be successful in a 21st century economy.

6. Launch rigorous U.S. intervention research to systematically replicate the international correlation findings (e.g. the mathematics-science performance association). One priority is to replicate the observed international correlational evidence of an association between mathematics and science performance through rigorous experimental designs. For example, interventions could seek to strengthen U.S. students' mathematics knowledge and simultaneously incorporate more mathematics into science curricula. A second priority is to validate the merits of delaying the start of science instruction. Interventions might offer young children serious, inquiry based science instruction beginning with grade 3, while focusing on reading and mathematics in kindergarten through grade 2. Alternatively, research could assess the feasibility of delivering more meaningful science instruction from kindergarten or grade 1.

In conclusion, the current benchmarking studies provide a global laboratory of suggestive, empirically based strategies for strengthening U.S. science performance. The international comparisons that benchmark U.S. performance against a consistent set of industrialized nations contribute to our knowledge of ways to improve U.S. science performance. Our findings show that the U.S. ranking decline begins in middle school and that reform efforts should not wait until secondary school science.

REFERENCES

American Association for the Advancement of Science. (2004). About *Project 2061*. Retrieved August 2007 from http://www.project2061.org/about/default/htm

American Association of University Women Educational Foundation. (2000). *Tech-Savvy: Educating girls in the new computer age*. Washington, DC: Author.

Bransford, J., Brown, A., and Cocking, R. (Eds.) (2000). *How people learn: Brain, mind, experience, and school.* Committee on Developments in the Science of Learning. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Brunner, J. (1960). *The process of education.* Cambridge, MA: Harvard University Press.

California State Board of Education. (1998). California Science Content Standards. Available at http://www.cde.ca.gov/be/st/ss/scintro.asp

Cohen, D. K., & Hill, H. C. (2001). *Learning policy*. New Haven, CT: Yale University Press.

Darling-Hammond, L. (2000, May). How teacher education matters. *Journal of Teacher Education*, *51*(3), 166–173.

Ferguson, P., & Womack, S. T. (1993). The impact of subject matter and education coursework on teaching performance. *Journal of Teacher Education*, *44*, 155–163.

Ginsburg, A., Cooke, G., Leinwand, S., Noell, J., & Pollock, E. (2005). *Reassessing U.S. international mathematics performance: New findings from the 2003 TIMSS and PISA*. Washington, DC: American Institutes for Research. Available at http://www.air.org/news/documents/TIMSS_PISA%20math%20study.pdf

Machamer, P. (Ed.). (1998). *The Cambridge companion to Galileo.* Cambridge: Cambridge University Press.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*, 125–145.

Mullis, I., Martin, M., Smith, T., Garden, R., Gregory, K., Gonzalez, E., Chrostowski, S., & O'Connor, K. (2003). *TIMSS assessment frameworks and specifications 2003.* Chestnut Hill, MA: Boston College, Lynch School of Education, International Study Center, International Association for the Evaluation of Educational Achievement (IEA). Retrieved May 2005 from http://timss.bc.edu/timss2003i/PDF/t03_AF_preface.pdf.

Mullis, I., Martin, M., Gonzalez, E., & Chrostowski, S. (2004). *TIMSS 2003 international mathematics report.* Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center. Retrieved May 2005 from http://timss.bc.edu/PDF/t03_download/T03_M_Front.pdf

National Academy of Sciences. (2005). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology. Washington, DC: National Academy of Sciences, National Academy of Engineering, Institute of Medicine. Retrieved October 2005 from http://www.nap.edu/catalog/11463.html.

National Academy of Science (in progress). *Learning science in informal environments: A review of research past, present, and future.* Retrieved September 2007 from http://www8.nationalacademies.org/cp/projectview.aspx?key=47544

National Research Council. (1996). *National science education standards.* National Academy of Sciences. Retrieved August 2007 from http://www.nap.edu/readingroom/books/nses/notice.html

National Research Council. (2002). *Learning and understanding: Improving advanced study of mathematics and science in U.S. high schools.* Committee on Programs for Advanced Study of Mathematics and Science in American High Schools. J. P. Gollub, M. W. Bertenthal, J. B. Labov, & P.C. Curtis (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Science Foundation. (2004). *International study shows select improvements in U.S. math and science performance – gains also seen among some minority students* (Press Release 04-157). Retrieved October 4, 2007, from http://www.nsf.gov/news/news_summ.jsp?cntn_id=100594&org=OISE&from=news

OECD. (2004). *Learning for tomorrow's world – first results from PISA 2003.* Retrieved August 2007 from http://www.oecd.org/document/55/0,3343,en_32252351_32236173_33917303_1_1_1_1,00.html

O'Day, J. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. H. Fuhrman (Ed.), *Designing coherent education policy: Improving the system*. San Francisco: Jossey Bass. (ED 359 626)

Roth, Kathleen (2007). Science teaching around the world: lessons for U.S. classrooms. *Education week*. Oct.10, 2007, p.9.

Sadler, P., & Tai, R. (2007, July). The two high-school pillars supporting college science. *Science*, pp. 457–458.

Schmidt, W., & McKnight, C. (1998, December). What can we really learn from TIMSS? (Third International Mathematics and Science Study). *Science*, p. 1830.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. C., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.

Victoria Curriculum and Assessment Authority. (2002). Sc*ience curriculum and assessment framework.* Retrieved July 2007 from http://csf.vcaa.vic.edu.au/sc/kosc.htm.

Wilson, S., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations.* Seattle: University of Washington, Center for the Study of Teaching and Policy.

Wisconsin Department of Public Instruction. (2005). *Wisconsin model academic standards for science.* Retrieved July 2007 from http://dpi.state.wi.us/standards/sciintro.html

World Bank. (2006). World development indicators. Retrieved September 2007 from http://devdata.worldbank.org/wdi2006/contents/income.htm

APPENDIX:
PILOT SCIENCE STANDARDS COMPARISON:
U.S., JAPAN AND AUSTRALIA

This section complements the prior quantitative analyses from the three international assessments with a pilot qualitative examination of science standards that identify what students should know and be able to do in different countries. The qualitative analyses of standards may help explain the prior quantitative findings about the importance mathematics, curricula, and teacher factors.

A full examination of the science standards in each of the 12 industrialized countries would be a useful undertaking, but such an analysis is far beyond the scope of this report. Instead, we conducted a pilot analysis to examine the potential benefits of a standards comparison. The pilot compares the science standards in the United States with those of Japan, one of the top scorers on the three international science assessments, and Australia, a western country that improves its score across assessments as the U.S. score declines.

The standards comparison consists of two parts. First, a discussion of the general organization of the standards describes their specificity by grade, components, and the science content component. Second, the pilot looks at the treatment of electricity in detail. Electricity is a major physics topic, and a comparison of its treatment may help inform the reasons for the relatively poor U.S. performance in physics.

U.S. Science Standards

The United States has no formal national science standards; each state is responsible for developing its own science standards. It is beyond the scope of this report to examine the state standards, but the National Science Education Standards (NSES) developed by the National Academy of Sciences (1996) have served as benchmarks for many states (e.g., Massachusetts, Wisconsin) when developing their own standards. Accordingly, this pilot examination of U.S. standards focuses on the NSES.

Organization of the National Science Education Standards. When the NSES were initially developed in the mid-1990s, they were highly innovative in their effort to pull together the core of science education. The U.S. standards movement was only beginning to take hold (Smith & O’Day, 1993), and there were no federal accountability requirements like those in No Child Left Behind that required assessments in the core subjects, including science.

The NSES follow a grade-band structure similar to that of the 1989 National Council of Teachers of Mathematics standards. The NSES standards cover three broad grade band, K–4, 5–8, and 9–12. The grade bands specify what should be taught over a grade range but do not indicate sequencing within a grade and leave open the possibility of considerable repetition in the teaching of science grade by grade.

The organization of the NSES reflects a comprehensive strategic reform document that establishes standards for a wide range of science education, not only science content. The NSES document is organized by six standards categories presented in the following order:

- *The science teaching standards* describe what teachers of science at all grade levels should know and be able to do.
- *The professional development standards* present a vision for the development of professional knowledge and skill among teachers.
- *The assessment standards* provide criteria against which to judge the quality of assessment practices.
- *The science content standards* outline what students should know, understand, and be able to do in the natural sciences over the course of K–12 education
- *The science education program standards* describe the conditions necessary for quality school science programs.
- *The science education system standards* consist of criteria for judging the performance of the overall science education system.

Exhibit 8. NSES Electricity and Magnetism Standards (Grades K–8)

K–4
- Electricity in circuits can produce light, heat, sound, and magnetic effects. Electrical circuits require a complete loop through which an electrical current can pass.
- Magnets attract and repel each other and certain kinds of other materials.

5–8
- Electrical circuits provide a means of transferring electrical energy when heat, light, sound, and chemical changes are produced.

The logic of standards suggests a different ordering in which content standards for student learning come first and standards for teaching, professional development, and assessment follow and are aligned with the content standards. An unintentional effect of having six different standards with content standards in the middle may be to diminish the focus on the content standards and reflect a tentativeness at the time to develop national science content for fear of a backlash.

The eight NSES content standards are more inclusive than the content standards that form the basis for the TIMSS and PISA international assessments, which focus on traditional science content strands. Instead, the NSES content standards cover eight categories: unifying concepts and processes in science; science as inquiry; physical science; life science; earth and space science; science and technology; science in personal and social perspective; and history and nature of science. The NSES include the traditional science strands as three of the eight content components: life science, physical science, and earth and space science. In effect, the NSES dilute the focus on science content by adding five other standards: unifying concepts, science as inquiry, science and technology, science in personal and social perspective, and history and nature of science.

NSES electricity and magnetism standards at elementary and middle grades. The NSES for electricity are also general (Exhibit 8). The introductory K–4 electricity and magnetism standard does focus on the important idea that electricity produces light, heat, sound, and magnetic effects; that electrical circuits require a closed loop; and that magnets attract and repel certain kinds of material. The K–4 topics seem to be appropriately introductory. However, the 5–8 electricity and magnetism standard simply states that students should learn that “electrical circuits provide a means for transferring electrical energy when heat, light, sound, and chemical changes are produced.” This is far too general and says little more than the K–4 standard. No guidance is provided about the treatment of important electricity topics such as parallel or series circuits or resistance and voltage. In fact, circuits are not even mentioned, and the relationship between electricity and magnetism is not developed. Overall, the NSES treatment of electricity lacks specificity and depth.

Japanese Science Standards

Organization of standards. The organization of the science standards for high-performing Japan is unlike that of the U.S. standards. First, the Japanese science standards do not start until grade 3; the U.S. standards start at kindergarten. Japan elects to place greater emphasis on learning language and mathematics in the earliest grades, saving science for a more in-depth treatment in the later grades.[6] The fact that the Japanese readily catch up in science and Japan becomes the top scorer among our 12-country sample may suggest that little real science content is taught in kindergarten and grades 1 and 2 in the United States.

Unlike the grade bands the NSES employ, the Japanese elementary standards identify specific science topics in each elementary grade, 3–6. The Japanese standards shift to grade bands only for the lower secondary standards (7–9) and the upper secondary standards (10–12).

Third, the organization of the Japanese standards is around the core science content strands, unlike the U.S. standards in which science content is only one of six objectives. The Japanese elementary (3–6) science standards are organized into three content strands: living things and their environment, matter and energy, and earth and space.

The Japanese lower secondary school content standards (7–9) are divided into two strands: matter and energy and living creatures and natural phenomena, with instructional time equally divided between the two. These Japanese standards do not ignore topics such as inquiry as science, but unlike U.S. standards that treat these as separate topics, the Japanese standards integrate these cross-cutting topics into the two content strands. As an example, the matter and energy strand starts with an inquiry focus that has students explaining physical phenomena by employing “observations and experiments (i.e., inquiry-based science) to understand the regularity of light and sound and the properties of force and … to establish in a scientific way connections between these phenomena and everyday life.”

[6] High-scoring Singapore also does not start teaching science until grade 3. Singapore scored highest on TIMSS-4 and TIMSS-8 but is not included in this study because Singapore does not participate in PISA.

Exhibit 9. Japan’s Electricity and Magnetism Science Education Standards (grades 3–9)

Grade 3:
By attaching a miniature light bulb to a battery, investigation of the connecting path and the objects through which the electricity passes, enabling children to learn to think about electric circuits. Children should:

a) understand that there are different ways of connecting up objects so that electricity either passes through them or does not pass through them.
b) understand that there are substances which conduct electricity and others which do not.

By the use of a magnet, carrying out experiments with magnetic attraction and attaching objects to the magnet, thus enabling children to learn to think about the nature of magnetism. Children should:

a) understand that there are some substances that are attracted by magnets and some that are not. In addition, they should understand that there are some substances which, after being attached to a magnet, become magnetic themselves.
b) understand that different magnetic poles attract, and similar poles repel.

Grade 4:
(3) Investigating movements when a motor is attached to a storage battery and a photo cell, thus being enabled to think about the functioning of electricity. Children should understand that:

a) by changing the number of batteries or the order in which they are connected, the brightness of a bulb or the revolutions of a motor will change;
b) a motor can be driven by a photo cell.

Grade 5:
Electricity not covered
Grade 6:
(3) Causing an electric current to flow by the use of electromagnetism, and investigating changes in the magnetic strength, leading to children being enabled to think about the functioning of electric current. Children should understand that:

a) when an electric current is passed through a wire coil wound around an iron core, the iron core becomes magnetized, and when the direction of the current changes, the magnetic polarity changes.
b) the strength of an electromagnet varies according to the strength of the electric current and according to the number of turns in the electric wire coil.

Grades 7–9 (Lower Secondary)
Electric current and its uses. Observations and experiments concerned with electric circuits, with the aim of enabling students to understand the relationship between current and voltage as well as the movement of electric current, and being brought to a rudimentary approach and consideration concerning electric current, while establishing links with everyday life.

a) With regard to electric current, students should be able:
(i) to identify the way in which static electricity occurs when two different kinds of substances are rubbed together, that a force operates between two charged objects even when a space separates them, and that there is a relationship between static electricity and electric current.
(ii) to construct an electric circuit and when carrying out experiments to measure the electric current and voltage in a circuit, to identify the regularity with regard to electric current running through each part of the circuit and voltage applied to each part
(iii) when carrying out experiments to measure the current and voltage applied to electric wire, to identify the connection between current and voltage, and at the same time, to understand that there is resistance in the metal wire.

b) With regard to the uses of electric current, students should be able:
(i) when observing a magnetic field caused by a magnet and electric wire, to understand that a magnetic field is expressed by means of lines of magnetic force, and to know that a magnetic field is created around a coil.
(ii) when carrying out experiments using a magnet and a coil, to identify the fact that when a current is passed through a coil in a magnetic field, a force is created, and that by moving the coil or the magnet, an electric current can be obtained.
(iii) when carrying out experiments designed to generate light or heat through the use of electric current, identify the fact that heat and light can be extracted from an electric current, and that the quantity of light or heat generated will vary according to differences in the electric power.

The upper secondary science standards cover a range of different science subjects with the purposes of cultivating abilities and attitudes about carrying out science observations and deepening scientific knowledge. Introductory secondary school courses cover how particular scientific phenomena in biology, physics, and earth and space science influence human well-being and lifestyles.

Electricity standards at elementary and middle grades. The Japanese coverage of the electricity and magnetism topics illustrates how their science standards build a coherent and deep treatment of a physics topic across the grades (Exhibit 9). Grade 3 uses examples to introduce electricity, conductivity and magnetic attraction and repulsion. Grade 4 presents different ways of generating electricity and the use of electricity to drive a motor. Grade 6 connects electricity with magnetism. Grades 7–9 (lower secondary) deepen knowledge with experiments into the relationship among electricity flow, voltage, and resistance; observations on the relationship between electricity and magnetism in generating electricity; and the relationship between electrical energy and heat and light energy.

Victoria, Australia's Science Standards

Australia, like the United States, does not have a national curriculum in science or other subjects. This study chose to examine the science curricula standards of Victoria (2002), which includes Melbourne. Victoria is one of Australia's largest states and has a Web site on which the Victoria science standards were readily accessible and clearly described.

Organization of standards. The Victoria science standards are organized differently than those in the United States or Japan are in that they cover six successive levels of student achievement:

- Level 1 - End of Preparatory Year Level 2 - End of Year 2
- Level 3 - End of Year 4
- Level 4 - End of Year 6
- Level 5 - End of Year 8
- Level 6 - End of Year 10.

At each level, the Victoria science standards are described in terms of three components:

- *Learning outcomes* answer the question, What should students know and be able to do as an outcome of their learning at this level?
- The *curriculum focus* outlines the major content to be covered to implement instructionally the learning outcomes.
- The *indicators* answer the question, How do we know that students have achieved the learning outcomes?

Victoria's learning outcomes describe Australia's content standards. At levels 1 and 2, the standards cover very introductory science material, including introductions to the language of science and simple patterns in scientific observations (some objects are attracted to a magnet and others are not).

Exhibit 10. Science Education Standards (Levels 1–5), Victoria, Australia

Identify transformations of energy involving electricity, light, sound, heat and movement
Level 3:
- identify light, sound, electricity, heat and movement as forms of energy
- identify energy changes that take place in common appliances
- demonstrate transformations of energy from electrical to light, heat, movement and sound in simple electric devices.

Level 4
- design, build and describe the operation of simple devices that transfer or transform energy.(investigate the operation of an electric torch to note the battery, globe, switch and connections)

Level 5
Describe the operation of direct current (DC) series and parallel circuits in terms of current and voltage.
This is evident when the student is able to:
- correctly connect commonly used components in series and parallel circuits
- describe the operation of series and parallel circuits, using terms such as current and voltage
- identify from circuit diagrams those circuit elements that are connected in series and those that are connected in parallel
- relate the brightness of a torch globe to the magnitude of the current in it and the voltage across it
- link the brightness of two or more globes connected in series and in parallel circuits to the magnitude of the voltage and current.

Identify the action of forces in everyday situations
Level 3
- identify the attraction/repulsion forces between the ends of two bar magnets

Level 4
- Describe the motion of objects in terms of simple combinations of forces

Level 5
Describe simple magnetic and electrostatic effects in terms of a field model.
This is evident when the student is able to:
- describe field patterns surrounding differently shaped magnets and simple combinations of magnets
- describe attraction and repulsion of magnets and objects near magnets as effects of the magnetic field
- explain the action of a compass as the movement of a magnet in the magnetic field of the Earth
- explain attraction and repulsion of charged objects in terms of the electric field around them
- describe the behavior of objects in an electric field.

Beginning in grade 3, the science standards are broken down into four strands: biological, chemical, earth and space, and physical sciences. The Australian standards achieve across-grade coherence for levels 3–6 by focusing consistently on building student science understanding within two broad learning areas for each strand across grades:

- *Biological science*: 1. Living together: past, present and future; 2. Structure and function

- *Chemical science:* 1. Substances: structure, properties and uses; 2. Reaction and change

- *Earth and space sciences*: 1. The changing Earth; 2. Our place in space
- *Physical science:* 1. Energy and its uses; 2. Forces and their effects

Electricity standards at elementary and middle grades (levels 1–5). Like the Japanese standards, the Australian electricity standards describe specific electricity topics that clearly build knowledge about the nature, workings, and uses of electricity from level to level (Exhibit 10). Australia does not formally introduce electricity as a distinct topic until level 3 (grade 4), at which time electricity is taught as a form of energy. Level 4 explains the applications of electricity concretely in the form of simple electrical devices. Level 5 examines the workings of electrical circuits in some detail, including parallel circuits and the relationship between current and voltage.

Summary

The U.S. National Academy of Sciences developed the NSES during the mid-nineties when the standards movement was beginning to take hold. Given the history of decentralized responsibility for education decisions, the general nature of the content standards with their grade bands and lack of topic development is not surprising. The standards from a consistently high-performing country such as Japan and those of Australia, whose scores improve to surpass the United States in secondary school, suggest that the NSES need to be reviewed and revised to offer coherent and more specific guidance on science content that is presented over narrower grade ranges.