OREA
Offices of Research and Education Accountability
Tennessee Comptroller of the Treasury, Justin P. Wilson
Legislative Brief

International Assessments

Kim Potts, Principal Legislative Research Analyst
(615) 401-7875/ Kim.Potts@tn.gov

April 2011

Key Points

The United States participates in three major international assessments that are designed to provide information about the performance of the U.S. K-12 education system relative to education systems in other countries:

- the Progress in International Reading Literacy Study (PIRLS);
- the Trends in International Mathematics and Science Study (TIMSS); and
- the Program for International Student Assessment (PISA).

The three assessments differ in purpose, subjects tested, and grade (or age) of students tested.

Individual countries' rankings on international tests do not tell the whole story. The rankings that are most frequently reported in news stories are usually based on overall averages, which can mask significant variations among states and subpopulations of test takers. Data from both TIMSS and PISA, for example, show that U.S. schools with lower percentages of socioeconomically disadvantaged students have average scores comparable to or higher than the highest-performing education systems.

Information gleaned from international test results has contributed to U.S. education policy over the years. For example, information from TIMSS about other countries' mathematics curriculum informed a lengthy national discussion about mathematics curriculum in the U.S., which is popularly described as "a mile wide and inch deep." The recent common core standards project represents an attempt to address the issue. The testing organizations that administer the international assessments publish a variety of supplementary materials that can be used to identify other countries' promising practices and policies and to analyze assessment data in greater depth.

Results from international tests should be interpreted with caution. Policymakers and educators in states or school districts in the U.S. may choose to make curriculum changes or alter graduation requirements based on findings from international assessments, but these actions cannot guarantee improvement in student achievement. International testing organizations caution against broadly interpreting international test results without placing other countries' education systems in context.

Efforts are underway to assist states with international benchmarking. The National Center for Education Statistics is conducting a study in 2011 that will link national and international assessments. When implemented, individual states will be able to measure the U.S. National Assessment of Educational Progress (NAEP) student performance results against international benchmarks as defined by TIMSS. The goal of the study is to predict 2011 TIMSS mathematics and science scores at grade 8 based on their NAEP performance for states, without incurring the costs associated with every state participating in TIMSS.

Introduction

The United States participates in three major international assessments designed to provide information about the performance of the U.S. K-12 education system relative to education systems in other countries: the Progress in International Reading Literacy Study (PIRLS); the Trends in International Mathematics and Science Study (TIMSS); and the Program for International Student Assessment (PISA). (See Exhibit 1 for more information about each test.)

The three assessments differ in purpose, subjects tested, and grade (or age) of students tested. Both PIRLS and TIMSS are designed to measure how well students are learning what they are taught in the classroom—PIRLS tests 4th grade reading every five years and TIMSS tests 4th and 8th grade mathematics and science every four years.[1] PISA is designed to measure whether 15-year-olds are able to practically apply what they have learned both in and out of school in reading, mathematics, and science; the test is administered every three years. PISA tests students in all three subjects, but focuses on one of the three subjects each time the test is given. For example, the 2006 PISA focused on science, the 2009 PISA focused on reading, and the 2012 PISA will focus on mathematics.

> To help countries make informed decisions about reading education, IEA's Progress in International Reading Literacy Study (PIRLS) provides internationally comparative data about students' reading achievement in primary school (the fourth grade in most participating countries). The fourth grade is an important transition point in children's development as readers, because most of them should have learned to read, and are now reading to learn.[2]
>
> TIMSS is designed to align broadly with mathematics and science curricula in the participating countries. The results, therefore, suggest the degree to which students have learned mathematics and science concepts and skills likely to have been taught in school.[3]
>
> PISA seeks to measure how well young adults, at age 15 and therefore approaching the end of compulsory schooling, are prepared to meet the challenges of today's knowledge societies—what PISA refers to as 'literacy.' The assessment is forward looking, focusing on young people's ability to use their knowledge and skills to meet real-life challenges, rather than merely on the extent to which they have mastered a specific school curriculum.[4] A country's results in PISA depend on the quality of care and stimulation provided to children during infancy and the pre-school years, and on the opportunities children have to learn both in school and at home during the elementary and secondary school years.[5]

The National Center for Education Statistics (NCES) sponsors all three assessments in the U.S. Two of the tests—PIRLS and TIMSS—are products of the International Association for the Evaluation of Educational Achievement (IEA). PISA is a product of the Organization for Economic Co-operation and Development (OECD). (See the shaded box on page 7 for more about the IEA and the OECD.)

The purpose of this brief is to describe:

- the international K-12 assessments in which the U.S. participates,
- the uses and limitations of international assessments,
- how the U.S. National Assessment of Educational Progress (NAEP) compares to the international assessments, and
- major results of the international assessments, with particular attention to U.S. results.

See Appendix A for a list of participating jurisdictions in each international assessment by year. See Appendix B for a description of each international assessment's benchmarks. See Appendix C for a select list of resources, including reports from NCES, IEA, and OECD.

Exhibit 1: U.S. Participation in International Assessments

	PIRLS (Progress in International Reading Literacy Study)	TIMSS (Trends in International Mathematics and Science Study)	PISA (Program for International Student Assessment)
What year did the study begin?	2001	1995	2000
How often is the study conducted?	Every 5 years	Every 4 years	Every 3 years
When will the study be conducted next?	2011	2011	2012
How many jurisdictions participated in the last assessment? How many were OECD countries?*	45 jurisdictions (19 OECD countries)	58 jurisdictions (Grade 4: 15 OECD countries; Grade 8: 11 OECD countries)	57 jurisdictions (30 OECD countries)
What is the target population?	Fourth-graders	Fourth- and eighth-graders	15-year-olds
How many U.S. participants were in the most recent study?	5,190	Grade 4: 10,350 Grade 8: 9,723	5,611
What is assessed?	Reading literacy	Mathematics, science	Reading, mathematical, and scientific literacy, with one subject assessed in depth at each administration (on a rotating basis) and the other two subjects as minor domains
What is the purpose of the study?	PIRLS measures students' reading comprehension of literary and informational text, broadly aligned with curricula of the participating countries.	TIMSS measures the mathematics and science knowledge and skills broadly aligned with curricula of the participating countries.	PISA measures how well students can apply their knowledge and skills to problems within real-life contexts. PISA is designed to represent a "yield" of learning at age 15, rather than a direct measure of attained curriculum knowledge.
Are descriptions provided of what the participants know and can do at various levels of performance?	Yes, international benchmarks at Advanced, High, Intermediate, and Low levels include descriptions of typical knowledge and skills.	Yes, international benchmarks at Advanced, High, Intermediate, and Low levels include descriptions of typical knowledge and skills.	Yes, Levels 1 (lowest) through 6 (highest) include descriptions of typical tasks that can be completed.

	PIRLS (Progress in International Reading Literacy Study)	TIMSS (Trends in International Mathematics and Science Study)	PISA (Program for International Student Assessment)
What scale scores are provided?	Reading literacy: Overall scale score and subscale scores	Grade 4 and Grade 8 Mathematics: Overall scale score and subscale scores Science: Overall scale score and subscale scores	Reading literacy: Overall scale score and subscale scores Mathematics literacy: Overall scale score and subscale scores Science literacy: Overall scale score and subscale scores
Are there sources that provide contextual information for the scale scores?	Yes, there are student, teacher, and school questionnaires, and in most countries (not including the United States) a parent questionnaire; national research coordinators also report on characteristics of national curriculum and selected education policies and practices.	Yes, there are student, teacher, and school questionnaires; national research coordinators also report on characteristics of national curriculum and selected education policies and practices.	Yes, there are student and school questionnaires, and in some countries (not including the United States) a parent questionnaire.

Notes: * There are a total of 30 countries in the Organization for Economic Cooperation and Development (OECD). The count for the OECD countries includes single countries and jurisdictions of the country (e.g., England and Scotland as representing the United Kingdom).
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, International Activities Program, Excerpted from table: "U.S. Participation in International Assessments," http://nces.ed.gov (accessed Jan. 10, 2011).

Uses and limitations of international assessments

Individual countries' rankings on international tests do not tell the whole story. News stories about international assessments frequently focus on how the U.S. ranks compared to other countries, but researchers note that such "horse race" comparisons should be made with caution. The rankings that are most frequently reported are usually based on overall averages, which can mask significant variations among states and subpopulations of test takers. The OECD notes that there is considerable performance variability in the U.S., which becomes apparent when the PISA 2009 data is disaggregated by groups of states:

> Such a comparison suggests that in reading, public schools in the northeast of the United States would perform at 510 PISA score points – 17 score points above the OECD average (comparable with the performance of the Netherlands) but still well below the [highest-performing education systems, including the Republic of Korea, Finland, Canada, New Zealand, and Japan] – followed by the midwest with 500 score points (comparable with the performance of Poland), the west with 486 score points (comparable with the performance of Italy) and the south with 483 score points (comparable with the performance of Greece)…[6]

In another example, an OECD and NCES analysis found that U.S. schools with lower levels of socioeconomically disadvantaged students have average PISA 2009 scores comparable to or higher than the highest-performing education systems, as shown in Exhibit 2. (An IEA and NCES analysis resulted in similar conclusions regarding the TIMSS 2007 average mathematics and science scores.)[7]

Results and findings from international tests have contributed to U.S. education policy over the years. NCES suggests that the most important uses for international assessment data are identifying promising practices and policies in other countries and considering patterns of student achievement in an international context, "including comparisons of average achievement and performance at the high and low ends of the distribution within and across countries."[8] Both the IEA and the OECD publish a variety of materials that supplement the test results for TIMSS, PIRLS, and PISA, and which can be used to identify other countries' promising practices and policies and to analyze the data in greater depth.

Example #1: Differences in curriculum

Data from the TIMSS assessment given in 1995 informed weaknesses in the U.S. approach to mathematics instruction, and may have prompted some states to consider curriculum changes. The 1995 TIMSS included an analysis of 491 curriculum guides and 628 textbooks from countries around the world, as well as data on teacher practices from the U.S., Germany, and Japan. Several subsequent reports dissected the analysis and found significant differences between the U.S. approach to teaching mathematics and the approaches used by the highest-performing countries. In brief, these reports found the U.S. mathematics curriculum to be "a mile wide and an inch deep" and "unfocused in comparison with those in other countries studied." The U.S. mathematics curriculum, which researchers found varied little among states, covered many more topics in much less depth and repeated many topics in several grades.[9] The common core standards movement emerged following this long-time concern about the U.S. curriculum.

> The CCSS (Common Core State Standards) bring a new focus and coherence to the mathematics curriculum. These standards avoid the "mile wide and inch deep" problem that has characterized American education.[10]

Example #2: Differences in number of math and science courses taken in high school

Because TIMSS 1995 included an assessment of students in their final year of secondary schooling (12th grade in the U.S.), the test results also made U.S. educators and policymakers aware that "U.S. graduating students were less likely to be taking mathematics or science than were their counterparts in other countries."[11] In recent years, many states (including Tennessee) have revised graduation requirements to include an additional year of mathematics and science.

Example #3: Students scoring at the advanced level in mathematics

With the recent release of the PISA 2009 results, some researchers have been concerned about the low percentage of U.S. students scoring at the advanced level in mathematics compared to the highest scoring countries. PISA describes six mathematics literacy proficiency levels ranging from level 1 to level 6, the most advanced. In the 2009 PISA assessment, 27 percent of U.S. students scored at or above proficiency level 4. This is lower than the 32 percent of students in

Exhibit 2: Average scores of U.S. 15-year-old students on combined reading literacy scale, by percentage of students in public school eligible for free or reduced-price lunch: PISA 2009

Percent of students eligible for free or reduced lunch	Average score on PISA 2009
Less than 10	551
10 to 24.9	527
25 to 49.9	502
U.S. average	*500*
OECD average	*493*
50 to 74.9	471
75 or more	446

Notes: (1) U.S. average score = 500. OECD average score = 493. Highest scoring countries: Republic of Korea (539); Finland (536); Canada (524); New Zealand (521); Japan (521); Australia (515).
(2) The National School Lunch Program provides free or reduced-price lunch (FRPL) for students meeting certain income guidelines. The percentage of students receiving FRPL is an indicator of the socioeconomic level of families served by the school.
(3) The OECD average is the average of the national averages of the OECD member countries, with each country weighted equally.
(4) Data are for public schools only.
Source: H.L. Fleischman, P.J. Hopstock, M.P. Pelczar, and B.E. Shelley, *Highlights from PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context*, NCES 2011-004, U.S. Department of Education, National Center for Education Statistics, 2010, p. 15, http://nces.ed.gov (accessed Jan. 25, 2011).

the OECD countries on average that scored at or above level 4.[12] At level 4 students can "complete higher order tasks" such as "solving problems that involve visual or spatial reasoning...in unfamiliar contexts."

On the other end of the spectrum, 23 percent of U.S. students scored below level 2, similar to the OECD average of 22 percent. Below level 2 students may not be able to consistently "employ basic algorithms," or make "literal interpretations of the results" of mathematical operations in real-life settings.[13] Comparing the data at the highest proficiency level across countries provides another, perhaps more concerning picture. The U.S. has only 1.9 percent of students scoring at the most advanced level on PISA, level 6, compared to Shanghai China with 26.6 percent of students scoring at level 6, Singapore with 15.6 percent, Chinese Taipei at 11.3 percent, and Hong Kong at 10.8 percent. Many other countries share a similar concern, as the OECD average for level 6 is 3.1 percent.[14]

Results from international tests should be interpreted with caution. These findings and the relative ranking of the U.S. regions and other groups of students point to significant differences in the education systems within the U.S. and among all participating countries, but are unable to supply information about what the differences are or what caused them. Policymakers and educators in states or school districts in the U.S. may choose to make curriculum changes or alter graduation requirements based on findings from international assessments, but these actions cannot guarantee improvement in student achievement. Both the IEA and the OECD caution against broadly interpreting international test results without placing other countries' education systems in context.[15]

According to NCES,

> Cultural and policy settings differ so extensively across countries that international assessments have limited utility for drawing conclusions about influences on student achievement. For example, what "school autonomy" means in the national education systems of most countries may be quite different from what it means in the United States where education is delivered through more than 13,000 school districts (an education organization that does not exist in most of the rest of the world). Further, statistical analyses that seek to identify which policies are associated with higher national performance on international assessments are, in the end, correlations between answers to questionnaires filled out by government officials and test scores. Correlation is not causation even under the best of analytic circumstances.[16]

Relationship between NAEP and international assessments

Under the auspices of the National Center for Education Statistics, the first national assessments of the U.S. K-12 educational system—the National Assessment of Educational Progress (NAEP)—were held in 1969. The first voluntary national assessments for states were held in the 1990s on a trial basis; these became a permanent feature of NAEP every two years. More recently, under No Child Left Behind, states receiving Title I grants are required to participate biennially in NAEP state-level assessments for reading and mathematics at grades 4 and 8.

NAEP is the largest nationally representative and continuing assessment of what U.S. students know and can do in various subject areas. Assessments are conducted periodically in mathematics, reading, science, writing, the arts, civics, economics, geography, and U.S. history. Since NAEP assessments are administered uniformly using the same sets of test booklets across the nation, NAEP results serve as a common metric for all states and selected urban districts. The assessment stays essentially the same from year to year, allowing NAEP to provide a clear picture of student academic progress over time.[17]

Differences between NAEP and international tests

The international assessments differ from the National Assessment of Educational Progress (NAEP) in several important ways:[18]

- Goals: NAEP is designed to measure the knowledge, skills, and competencies needed by U.S. students; PIRLS, TIMSS, and PISA are each developed in an internationally collaborative manner to reflect the interests of a wide range of countries, including the U.S.

- Participating countries: NAEP is given only in the U.S. The IEA (which administers TIMSS and PIRLS) includes in its assessments a diverse group of countries and jurisdictions, some of which are developing countries. OECD assessments are conducted in all 30 member nations of the OECD, with some participation by non-OECD countries and jurisdictions.

- Populations tested: NAEP tests students in grades 4, 8, and 12; PIRLS tests only grade 4; TIMSS tests grades 4 and 8 (and has tested grade 12 twice); and PISA tests 15-year-olds.

- Sample sizes on which the tests are based: All of the assessments are sample-based. Because NAEP is reported at the state level, NAEP samples a much larger number of students than does PIRLS, TIMSS, and PISA. (See Exhibit 3.) The NAEP national sample comprises individual state samples of public school students, supplemented by a national sample of nonpublic school students. According to NCES, "NAEP generally measures performance at a finer level of precision than TIMSS or PISA, and these differences can have an impact on the assessments' sensitivities in detecting changes in student performance."[19]

- Content: According to NCES, NAEP and the international assessments differ (sometimes considerably) in various aspects of the content tested, including content coverage and item format. "Overall assessment scores can depend on the extent to which the proportion of the items devoted to various topics or skills aligns with the emphases of the education system's curriculum."[20] In the mathematics assessments, TIMSS and NAEP both have a majority of multiple-choice test items, and about two-thirds of the items on PISA are constructed response. Item formats in science, mostly multiple choice and some constructed response, are similar on all three tests.

IEA and OECD: International Testing Organizations

The IEA is an independent, international cooperative of national research institutions and governmental research agencies. Overall policy for TIMSS and PIRLS is set through meetings of the IEA General Assembly with representatives from member countries. Operational decisions are made through meetings of National Research Coordinators and representatives of the IEA and the International Study Center at Boston College, which has the international contract for developing and administering TIMSS and PIRLS.

The OECD is an intergovernmental treaty organization composed mainly of industrialized countries. Overall policy for PISA is set through meetings of the PISA Governing Board, which consists of representatives from the 30 OECD countries; non-OECD countries participating in PISA may also send representatives to these meetings as observers. Operational decisions are made through meetings of National Program Managers and representatives of the OECD and the Australian Council for Educational Research (ACER) and other members of the consortium of contractors, which has the international contract for developing and administering PISA.

Within the United States, TIMSS, PIRLS, and PISA are sponsored by NCES. NCES staff act as U.S. National Research Coordinators for TIMSS and PIRLS and as the U.S. National Program Manager for PISA. NCES staff oversee the work of national data collection contractors, which implement the three studies in the United States.

Source: Mark Schneider, Commissioner, NCES, "Benefits and Limitations of States Benchmarking to International Standards: A Meeting to Assist States in Making Informed Decisions About Participating in International Assessments," May 30, 2008, http://nces.ed.gov (accessed March 7, 2011).

Linking NAEP results to the results of international assessments

Recent U.S. news reports attest to widespread state interest in "international benchmarking," a phrase that generally refers to comparing U.S. student performance to the highest performing educational systems in the world. In 2007, in addition to the U.S. participating in TIMSS as a nation, two states—Massachusetts and Minnesota—opted to participate as "benchmarking entities." (Other benchmarking entities included in the 2007 TIMSS were the Canadian provinces of Alberta, British Columbia, Ontario, and Quebec; Dubai (United Arab Emirates); and the Basque region of Spain.)[21] The cost of participation in international assessments is expensive. A former NCES commissioner has estimated the cost at $25 million if every state participated in PISA and TIMSS, for example.[22] To allow individual states to measure their student performance against international benchmarks, in 2011 NCES is conducting a new study to link NAEP and TIMSS.

Exhibit 3: Number of students and schools participating in NAEP, TIMSS, and PISA

	Number of students	Number of schools
NAEP 2007 Mathematics (4th grade)	197,700	7,840
NAEP 2007 Mathematics (8th grade)	153,000	6,910
NAEP 2005 Science (4th grade)	147,700	8,500
NAEP 2005 Science (8th grade)	143,400	6,400
NAEP 2005 Science (12th grade)	13,700	900
TIMSS 2007 (4th grade)	7,900	260
TIMSS 2007 (8th grade)	7,400	240
PISA 2006	5,600	170
PISA 2009	5,233	165

Note: Numbers have been rounded to the nearest hundred for students and the nearest 10 for schools.
Sources: National Center for Education Statistics, *The Nation's Report Card: Mathematics 2007* (NCES 2007-094); *The Nation's Report Card: Science 2005* (NCES 2006-466); *Highlights from TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009-001); *Highlights from PISA 2006: Performance of U.S. 15-Year-Old Students in Science and Mathematics Literacy in an International Context* (NCES 2008-016); and NCES, "Program for International Student Assessment (PISA)," Frequently Asked Questions http://nces.ed.gov/surveys/pisa/faq.asp.

Three previous attempts have been made to link NAEP and TIMSS[23] and one recent effort sought to link NAEP and PISA.[24] In the most recent effort to link 2007 NAEP and TIMSS data, Dr. Gary W. Phillips used "the method of statistical moderation by applying the correspondence between the national score distributions of NAEP and TIMSS to the state and district NAEP score distributions."[25] (See the box titled "States' results from 2007 report linking NAEP and TIMSS mathematics results" for more about Dr. Phillips' study.)

The NCES linking study will use a different approach. According to NCES:

> The NAEP-TIMSS Linking Study, to be conducted in the first half of 2011, will take a new approach to using grade 8 mathematics and science data from the National Assessment of Educational Progress (NAEP) to project state-level scores on the Trends in International Mathematics and Science Study (TIMSS). The goal of the study is to predict 2011 TIMSS mathematics and science scores at grade 8 based on their NAEP performance for states, without incurring the costs associated with every state participating in TIMSS.[26]

In the study, eight states will use actual state TIMSS results to validate the predicted TIMSS average scores. The eight states are Alabama, California, Colorado, Connecticut, Indiana, Massachusetts, Minnesota, and North Carolina.

States' results from 2007 report linking NAEP and TIMSS mathematics results

A 2007 study by Dr. Gary W. Phillips of the American Institutes for Research was able to "statistically link" NAEP and TIMSS, allowing researchers to estimate how each state's students would have performed on TIMSS based on how they had performed on NAEP. According to the study, Tennessee's 4th graders would have received a grade of C if the state had participated in TIMSS 2007. No state would have received an A and only five states (Massachusetts, Minnesota, New Jersey, New Hampshire, and Kansas) would receive a B—Vermont would receive a B-. Most states—36—would receive a C+; eight (including Tennessee) would receive a C; and one—Washington, D.C.—would receive a D+.

For the 8th grade, only one state (Massachusetts) received a B-; 24 states received a C+; 24 (including Tennessee) received a C; one received a C-; and one received a D+.

The study finds that there is a set of states (Massachusetts, Minnesota, New Jersey, New Hampshire, Kansas, and Vermont) in which students are learning at the B or B- level—equivalent to what the author described as the "international benchmark" that is above the international average of mathematics learning, which he finds to be a C level in grade 4 and a D+ in grade 8. According to the study's author, that these states are performing above the international average indicates that "it is possible in the United States for students to learn mathematics at a level that is competitive with the best in the world."

Source: G. W. Phillips, *The Second Derivative: International Benchmarks in Mathematics for American States and School Districts*, Washington, D.C., American Institutes for Research, 2009, http://www.air.org (accessed March 7, 2011).

Exhibit 4: Brief summaries of international assessment results with a focus on the U.S., by subject area and grade/age tested

The summaries that follow are primarily drawn from analyses done by the National Center for Education Statistics, the U.S. sponsor of PIRLS, TIMSS, and PISA. Information is supplied for the most recent assessment, along with brief comments about trends based on previous assessments. (See Appendix A for lists of the jurisdictions that participated in each assessment by year. See Appendix B for descriptions of the international benchmarks for PIRLS, TIMSS, and PISA.) Note: All international tests include participation by some political entities other than countries (such as Canadian provinces, separate nations that make up the United Kingdom, and administrative regions of China, for example). The term "country" is used in this table as a common name for the range of political entities that participated in each study.

Reading

- *The U.S. has participated in two assessments that measure reading: PIRLS and PISA.*
- *PIRLS assessed 4th graders in 2001 and 2006.*
- *PISA assessed 15-year-olds in 2000, 2003, 2006, and 2009.*

Grade/Age tested: 4th

Results:

Average Scores: 2006
On average, U.S. students in grade 4 scored higher than their peers worldwide, with average scores higher than the PIRLS scale average (540 vs. 500). The U.S. average score was higher than the average score for 22 of the 45 participating jurisdictions and lower than the average score for 10 jurisdictions. The U.S. average score was not measurably different from the remaining 12 jurisdictions (Austria, Belgium (Flemish), Bulgaria, Canada-Nova Scotia, Canada-Quebec, Chinese Taipei, Denmark, England, Germany, Latvia, Lithuania, and Netherlands).

Proficiency Levels: 2006
The U.S. had higher percentages of students than the international averages in each of the 4 international benchmarks for PIRLS 2006:

	% U.S. average	% international median
Advanced	12	7
High	47	41
Intermediate	82	76
Low	96	94

Trends
U.S. scores stayed about the same between 2001 and 2006, while the reading scores in 8 countries improved. In 2006, 10 countries outperformed the U.S. and the U.S. outperformed 22 countries. In 2001, 35 countries participated; in 2006, 45 countries participated.

Grade/Age tested:	**Results:**
15-year-olds	*Average Scores: 2009* U.S. 15-year-olds had an average score of 500 on the combined reading literacy scale, not measurably different from the average score of 493 for the 34 OECD countries. In 2009, among the 33 other OECD countries: • 6 countries outperformed the U.S. (Republic of Korea, Finland, Canada, New Zealand, Japan, and Australia), • 13 had lower average scores, and • 14 had average scores not measurably different from the U.S. average.

Among the 64 countries other than the U.S. taking the 2009 assessment (OECD countries, non-OECD countries, and other education systems):

- 9 had higher scores than the U.S. (the 6 listed above plus Shanghai-China, Hong Kong-China, and Singapore),
- 39 had lower average scores, and
- 16 had average scores that were not measurably different from the U.S. score.

Proficiency Levels: 2009
On the combined reading literacy scale, the U.S. had averages similar to the OECD average at every proficiency level:

	% U.S. average	**% OECD average**
Level 6 (most advanced level)	2	1
Level 5	8	7
Level 4	21	21
Level 3	28	29
Level 2	24	24
Level 1a	13	13
Level 1b	4	5
Below Level 1b	1	1

Trends
There was no measurable difference between the average score of U.S. students in reading literacy in 2000, the last time in which reading literacy was the major domain assessed in PISA, and 2009, or between 2003 and 2009. [Note: 2006 PISA reading literacy scores for the U.S. were not valid because of a printing error in the test booklets, so data cannot be compared for that assessment year in this subject.]

Mathematics

- *The U.S. has participated in two assessments that measure the mathematical skills of students: TIMSS and PISA.*
- *TIMSS assessed 4th graders in 1995, 2003, and 2007, and 8th graders in 1995, 1999, 2003, and 2007.*
- *PISA assessed 15-year-olds in 2000, 2003, 2006, and 2009.*

Grade/Age tested:	Results:
4th	*Average Scores: 2007*

In 2007, the average score of U.S. 4th-graders was 529 compared with the TIMSS scale average of 500. This was higher than averages in 23 of the 35 other countries taking the assessment, lower than those in 8 countries, and not measurably different from the average scores in the remaining 4 countries. A total of 36 countries participated in the 4th grade mathematics assessment in TIMSS 2007.

Proficiency Levels: 2007

The U.S. had higher percentages of students than the international averages in each of the four international benchmarks on TIMSS 2007:

	% U.S. average	% international median
Advanced	10	5
High	40	26
Intermediate	77	67
Low	95	90

Seven jurisdictions had higher percentages of students reaching the advanced proficiency level than the U.S.: Singapore (41 percent); Hong Kong SAR (40 percent); Chinese Taipei (24 percent); Japan (23 percent); Kazakhstan (19 percent); England (16 percent); and the Russian Federation (16 percent).

Highest and lowest performing students

To examine the mathematics performance of each participating country's higher and lower performing students, cutpoint scores were calculated for students performing at or above the 90th percentile (that is, the top 10 percent of students) and those performing at or below the 10th percentile (the bottom 10 percent of students). The cutpoint scores were calculated for each country, rather than across all countries combined.

In 2007, the highest-performing U.S. fourth-graders (those performing at or above the 90th percentile) scored 625 or higher. This was higher than the 90th percentile scores for 4th-graders in 23 countries and lower than the 90th percentile score for students in 7 countries. The countries in which the 90th percentile cutpoint score was higher than the cutpoint score for the U.S. are the same as those that outperformed the United States as a whole, with the exception of Latvia where the 90th percentile score of 628 is not significantly different from 625 in the United States. The 90th percentile scores ranged between 371 (Yemen) and 702 (Singapore). The difference in the 90th percentile score between Singapore, the highest performing country, and the United States was 77 score points.

The lowest-performing U.S. fourth-graders (those performing at or below the 10th percentile) scored 430 or lower in 2007. This was higher than the 10th percentile score in 23 countries and lower than the 10th percentile score in 6 countries: Singapore, Hong Kong SAR, Japan, Chinese Taipei, Latvia, and the Netherlands. The score at the 10th percentile ranged between 81 (Yemen) and 520 (Hong Kong SAR). The difference in the cutpoint scores between the lowest-performing students in Hong Kong SAR and the United States was 90 score points.

Trends
At grade 4, 16 countries, including the United States, participated in both the first and most recent TIMSS administrations. Between 1995 and 2007, one-half of the countries (8 of 16), including the United States, showed improvement in average scores and one-quarter of the countries (4 of 16) showed declines. In 2007, the U.S. 4th-grade average mathematics score of 529 was 11 scale score points higher than the 1995 average of 518.

The gain in the U.S. 4th-grade average mathematics score (11 scale score points) was greater than the difference in 6 countries (the 4 countries with declines in average scores, as well as 2 other countries) and less than the gain of 4 countries (England, Hong Kong SAR, Slovenia, and Latvia). There was no measurable difference between the 11 score point gain in the United States and the gains or declines in score points experienced in the other countries.

Grade/Age tested:

8th

Results:

Average Scores: 2007
In 2007, the average score of U.S. 8th-graders was 508 compared with the TIMSS scale average of 500. This was higher than averages in 37 of the 47 other countries, lower than those in 5 countries (all located in Asia), and not measurably different from the average scores in the remaining 5 countries. A total of 48 countries participated in the 8th grade mathematics assessment of TIMSS 2007.

Proficiency Levels: 2007
The U.S. had higher percentages of students than the international averages in each of the 4 international benchmarks for TIMSS 2007:

	% U.S. average	% international median
Advanced	6	2
High	31	15
Intermediate	67	46
Low	92	75

Seven jurisdictions had higher percentages of students reaching the advanced proficiency level than the U.S.: Chinese Taipei (45 percent); Republic of Korea (40 percent); Singapore (40 percent); Hong Kong SAR (31 percent); Japan (26 percent); Hungary (10 percent); and Russian Federation (8 percent).

Highest and lowest performing students
To examine the mathematics performance of each participating country's higher and lower performing students, cutpoint scores were calculated for students performing at or above the 90th percentile (that is, the top 10 percent of students) and those performing at or below the 10th percentile (the bottom 10 percent of students). The cutpoint scores were calculated for each country, rather than across all countries combined.

At grade 8, the highest-performing U.S. students (90th percentile or higher) in mathematics scored 607 or higher. The U.S. 90th percentile score was higher than that of 34 countries and lower than the 90th percentile score in 6 countries: Chinese Taipei, Korea, Singapore, Hong Kong SAR, Japan, and Hungary. The range at the 8th grade in 90th percentile scores was between 427 (Qatar) and 721 (Chinese Taipei). The difference in average scores between the 90th percentile in Chinese Taipei and the United States was 114 score points.

The lowest-performing U.S. 8th-graders (10th percentile or lower) scored 408 or less in 2007. The 10th percentile score for U.S. 8th-graders in mathematics was higher than the 10th percentile score in 34 countries and lower than the 10th percentile score in 4 countries: Chinese Taipei, Korea, Singapore, and Japan. The range in 10th percentile scores was between 186 (Qatar) and 475 (Korea). The difference in the cutpoint scores between the lowest-performing students in Korea and the United States was 66 score points.

Trends
At grade 8, 20 countries, including the United States, participated in TIMSS in both 1995 and 2007. About one-quarter of the countries (6 of 20), including the United States, had higher average mathematics scores in 2007 than in 1995 and students in one-half of the countries (10 of 20) showed declines in their average scores. The U.S. 8th-grade average mathematics score of 508 was 16 scale score points higher than the 1995 average of 492. The gain in the U.S. 8th-grade mathematics score (16 scale score points) was greater than the difference in 13 countries (including the 10 countries with declining scores and 3 others) and less than the gain of 2 countries (Colombia and Lithuania). There was no measurable difference between the 16 score point gain in the United States and the gains or declines in score points experienced in the other countries.

Grade/Age tested:

15-year-olds

Results:

Average Scores: 2009
The U.S. average score in mathematics literacy (487) was lower than the OECD average score (496) in 2009. Among the 33 other OECD countries, 17 countries had higher average scores than the United States, 5 had lower average scores, and 11 had average scores not measurably different from the U.S. average. The OECD countries with average scores higher than the U.S. average were: Korea, Finland, Switzerland, Japan, Canada, the Netherlands, New Zealand, Belgium, Australia, Germany, Estonia, Iceland, Denmark, Slovenia, Norway, France, and the Slovak Republic. The OECD countries with lower average scores than the United States were Greece, Israel, Turkey, Chile, and Mexico. Among all other 64 participating jurisdictions, including the OECD countries and non-OECD countries/jurisdictions, 23 had higher average scores than the United States, 29 had lower average scores, and 12 had average scores not measurably different from the U.S. average score. A total of 65 countries participated in the 2009 PISA (OECD countries, non-OECD countries, and other education systems).

Proficiency Levels: 2009
PISA describes 6 mathematics literacy proficiency levels ranging from level 1 to level 6, the most advanced. The percentage of U.S. students reaching the various levels was similar to or slightly above the OECD average on levels 1 through 4; the U.S. was slightly below the OECD average on levels 4 through 6, the most advanced levels:

	% U.S. average	% OECD average
Level 6 (most advanced level)	2	3
Level 5	8	10
Level 4	17	19
Level 3	25	24
Level 2	24	22
Level 1	15	14
Below Level 1	8	8

At level 4 students can "complete higher order tasks" such as "solving problems that involve visual or spatial reasoning...in unfamiliar contexts." Twenty-three percent of U.S. students scored below level 2. Below level 2 students may not be able to consistently "employ basic algorithms," or make "literal interpretations of the results" of mathematical operations in real-life settings.

Trends
The U.S. average score in mathematics literacy in 2009 was higher than the U.S. average in 2006 but not measurably different from the U.S. average in 2003, the earliest time point to which PISA 2009 performance can be compared in mathematics literacy.

Science

- *The U.S. has participated in two assessments that measure the science skills of students: TIMSS and PISA.*
- *TIMSS assessed 4th and 8th graders in 1995, 2003, and 2007.*
- *PISA assessed 15-year-olds in 2000, 2003, 2006, and 2009.*

Grade/Age tested:

4th

Results:

Average Scores: 2007

In 2007, the average score of 4th-graders was 539 compared to the TIMSS scale average of 500. At grade 4, the average U.S. science score was higher than those in 25 of the 35 other countries, lower than the average scores in 4 countries (all of them in Asia), and not measurably different from the average scores of students in the remaining 6 countries. A total of 36 countries participated in the 4th grade science assessment in TIMSS 2007.

Proficiency Levels: 2007

The U.S. had higher percentages of students than the international averages in each of the 4 international benchmarks for TIMSS 2007:

	% U.S. average	% international median
Advanced	15	7
High	47	34
Intermediate	78	74
Low	94	93

Two countries had higher percentages of students reaching the advanced proficiency level than the U.S.: Singapore (36 percent) and Chinese Taipei (19 percent).

Highest and Lowest Performing Students

To examine the science performance of each participating country's higher and lower performing students, cutpoint scores were calculated for students performing at or above the 90th percentile (that is, the top 10 percent of students) and those performing at or below the 10th percentile (the bottom 10 percent of students). The cutpoint scores were calculated for each country, rather than across all countries combined.

In 2007, the highest-performing U.S. 4th-graders (those performing at or above the 90th percentile) scored 643 or higher in science. This was higher than the 90th percentile score for 4th-graders in 27 countries and lower than 2 of the 35 other countries. Of the 4 countries that outperformed the United States, on average, in science at grade 4, 2 had higher 90th percentile cutpoint scores than the United States: Singapore and Chinese Taipei. Scores at the 90th percentile ranged between 379 (Yemen) and 701 (Singapore). The difference in scores between the highest-performing students in Singapore and the United States was 58 score points.

The lowest-performing U.S. 4th-graders in science (those performing at or below the 10th percentile) scored 427 or less in 2007. The 10th percentile score for U.S. 4th-graders was higher than the 10th percentile score in 17 countries and lower than that in 7 countries: Singapore, Chinese Taipei, the Russian Federation, Hong Kong SAR, Japan, Latvia, and the Netherlands. The range in scores at the 10th percentile was between 20 (Yemen) and 466 (Hong Kong SAR). The difference in scores between the lowest-performing students in Hong Kong SAR and the United States was 39 score points.

Trends
At grade 4, 16 countries, including the United States, participated in both the first TIMSS in 1995 and the most recent TIMSS in 2007 and therefore can be compared over a 12-year period. Comparing 2007 with 1995, 7 of the 16 countries showed improvement in average science scores, 5 countries showed declines and 4 countries, including the United States, had no measurable difference in average scores. In 2007, the U.S. 4th-grade average score was 539, compared with 542 in 1995.

A comparison of 1995 and 2007 shows a decline in the 90th percentile cut point score for U.S. 4th graders in science, the point marking the top 10 percent of students. In 2007, the 90th percentile score was 643, 11 score points lower than the analogous score of 654 in 1995. A comparison of the 10th percentile science scores for U.S. 4th-graders in 1995 and 2007 and 2003 and 2007 shows no measurable difference.

Grade/Age tested:

8th

Results:

Average Scores: 2007
In 2007, the average score of U.S. 8th-graders was 520, compared to the TIMSS scale average of 500. At grade 8, the average U.S. science score was higher than those in 35 of the 47 other countries, lower than in 9 countries (all located in Asia or Europe), and not measurably different from the average scores of students in the remaining 3 countries. A total of 48 countries participated in the 8th grade science assessment in TIMSS 2007.

Proficiency Levels: 2007
The U.S. had higher percentages of students than the international averages in each of the 4 international benchmarks for TIMSS 2007:

	% U.S. average	% international median
Advanced	10	3
High	38	17
Intermediate	71	49
Low	92	78

Six countries had higher percentages of students reaching the advanced proficiency level than the U.S.: Singapore (32 percent): Chinese Taipei (25 percent); Japan (17 percent); Republic of Korea (17 percent); and Hungary (13 percent)..

Highest and Lowest Performing Students
To examine the science performance of each participating country's higher and lower performing students, cutpoint scores were calculated for students performing at or above the 90th percentile (that is, the top 10 percent of students) and those performing at or below the 10th percentile (the bottom 10 percent of students). The cutpoint scores were calculated for each country, rather than across all countries combined.

The highest-performing U.S. 8th-grade students (90th percentile or higher) in science scored 623 or higher in 2007. This was higher than the 90th percentile score in 34 countries and lower than in 6 countries: Singapore, Chinese Taipei, England, Japan, Korea, and Hungary. The range in 90th percentile scores was between 445 (Ghana) and 694 (Singapore). The difference in scores between the highest-performing students in Singapore and the United States was 71 score points.

The lowest-performing U.S. 8th-graders (10th percentile or lower) scored 410 or lower in science in 2007. The 10th percentile score for U.S. 8th-graders was higher than the 10th percentile score in 34 countries and lower than in 8 countries: Chinese Taipei, England, Japan, Korea, Hungary, the Czech Republic, Slovenia, and the Russian Federation. The range in 10th percentile scores was between 163 (Ghana) and 454 (Japan). The difference in scores between the lowest-performing students in Japan and the United States was 44 score points.

Trends
At grade 8, 19 countries, including the United States participated in TIMSS in both 1995 and 2007. Five countries had higher average science scores in 2007 than in 1995, 3 countries showed declines in their average scores, and 11 countries, including the United States, had no measurable difference between average scores in 1995 and 2007. The U.S. 8^{th}-grade average science score was 520, compared with 513 in 1995.

At grade 8, the 90^{th} percentile cutpoint score in science showed no measurable differences in comparisons of 2007 to 1995 or 2003, but showed a decrease when the 2007 score was compared to the 1999 score (636 v. 623). The score identifying the lowest performing U.S. 8^{th}-graders in science was higher in 2007 than in 1995 (410 v. 384) and in 1999 (410 v. 386).

Grade/Age tested:

15-year-olds

Results:

Average Scores: 2009
On the science literacy scale, the average score of U.S. students (502) was not measurably different from the OECD average (501). Among the 33 other OECD countries, 12 had higher average scores than the United States, 9 had lower average scores, and 12 had average scores that were not measurably different from the U.S. average score. The OECD countries with higher average scores than the United States were: Finland, Japan, Korea, New Zealand, Canada, Estonia, Australia, the Netherlands, Germany, Switzerland, the United Kingdom, and Slovenia. The OECD countries with lower average scores than the United States were: the Slovak Republic, Italy, Spain, Luxembourg, Greece, Israel, Turkey, Chile, and Mexico. Among all other 64 participating jurisdictions, including the OECD countries and non-OECD countries/jurisdictions, 18 had higher average scores, 33 had lower average scores, and 13 had average scores that were not measurably different from the U.S. average score. A total of 65 countries participated in the 2009 PISA (OECD countries, non-OECD countries, and other education systems).

Proficiency Levels: 2009
PISA describes six science literacy proficiency levels ranging from level 1 to level 6, the most advanced. The U.S. was similar to the OECD average:

	% U.S. average	% OECD average
Level 6 (most advanced level)	1	1
Level 5	8	7
Level 4	20	21
Level 3	28	29
Level 2	25	24
Level 1	14	13
Below Level 1	4	5

Twenty-nine percent of U.S. students and students in the OECD countries on average scored at or above level 4 on the science literacy scale. At level 4 students “select and integrate explanations from different disciplines of science or technology” and “link those explanations directly to...life situations.” Eighteen percent of U.S. students and students in the OECD countries on average scored below level 2. Below level 2 students may not be able to consistently “provide... explanations in familiar contexts or draw conclusions based on simple investigations" or consistently "make literal interpretations.”

Trends
The U.S. average score in science literacy in 2009 was higher than the U.S. average in 2006. While U.S. students scored, on average, below the OECD average in science literacy in 2006, the average score of U.S. students in 2009 was not measurably different from the 2009 OECD average.

Sources: J. Baer, S. Baldi, K. Ayotte, and P. Green, *The Reading Literacy of U.S. Fourth-Grade Students in an International Context: Results From the 2001 and 2006 Progress in International Reading Literacy Study (PIRLS)* (NCES 2008–017), National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, D.C., 2007, http://nces.ed.gov; P. Gonzales, T. Williams, L. Jocelyn, S. Roey, D. Kastberg, and S. Brenwald, *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009–001 Revised), National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC, 2008, http://nces.ed.gov ; Howard L. Fleischman, Paul J. Hopstock, Marisa P. Pelczar, Brooke E. Shelley, *Highlights from PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context*, U.S. Department of Education, National Center for Educational Statistics, Washington, DC, U.S. Government Printing Office, NCES 2011-004, Dec. 2010, http://nces.ed.gov.

Notes:

1) PIRLS scores are reported on a scale from 0–1,000 with the scale average fixed at 500 and a standard deviation of 100. The PIRLS scale average was set in 2001 and reflects the combined proficiency distribution of all students in all jurisdictions participating in 2001. To allow comparisons between 2001 and 2006, scores of students in jurisdictions that participated in both 2001 and 2006 (29 jurisdictions) were used to scale the 2006 results. The 2006 scores were linked to the 2001 scale using common items on both assessments. Once scores from the 2006 assessment were scaled to the 2001 scale, scores of students in jurisdictions that participated in 2006 but not in 2001 were placed on the PIRLS scale.

   The PIRLS international benchmarks provide a way to interpret scale scores and to understand how students' proficiency varies along the PIRLS scale. In 2001, the cutpoints for the PIRLS benchmarks were set on the basis of the distribution of students along the PIRLS scale (the top 10 percent, the upper quartile, the median, and the lower quartile). In 2006, the cutpoints were revised to be identical to the cutpoints used for the Trends in International Mathematics and Science Study (TIMSS), which is also conducted by the IEA.

2) The total number of countries reported in the NCES *Highlights from TIMSS 2007* report differs from the total number reported in the international TIMSS reports. NCES did not include Morocco at the 8th grade level because of sampling difficulties noted by IEA. It did not include Mongolia at the 4th or 8th grade levels because the country could not complete the steps required to have its data included in the TIMSS report; the IEA's TIMSS 2007 report includes Mongolia's data in an appendix. NCES thus reports 36 participating countries at the 4th grade level; IEA reports 37. NCES reports 48 participating countries at the 8th grade level; IEA reports 50.

   In addition to the 36 countries at grade 4 and 48 countries at grade eight, 8 other educational jurisdictions, or "benchmarking" entities, participated in TIMSS 2007: the states of Massachusetts and Minnesota; the Canadian provinces of Alberta, British Columbia, Ontario, and Quebec; Dubai, United Arab Emirates; and the Basque region of Spain.

3) TIMSS provides two overall scales—mathematics and science—as well as several content and cognitive domain subscales for each of the overall scales. The scores are reported on a scale from 0 to 1,000, with the TIMSS scale average set at 500 and standard deviation set at 100. Differences are statistically significant at the .05 level.

   TIMSS reports on four benchmarks to describe student performance in mathematics and science. Each benchmark is associated with a score on the achievement scale and a description of the knowledge and skills demonstrated by students at that level of achievement. The advanced international benchmark indicates that students scored 625 or higher.

   According to NCES, all data presented in the TIMSS are used to describe relationships between variables. These data are not intended, nor can they be used, to imply causality. Student performance can be affected by a complex mix of educational and other factors that are not examined here.

For PISA, the OECD averages cited are the averages of the national averages of the OECD member countries, with each country weighted equally. The results for non-OECD countries are not included in the OECD average, though those countries' averages are included in the rankings of countries. Scores are reported on a scale from 0 to 1,000. Score differences between the United States and other countries (as well as between the U.S. and the OECD average) are significantly different at the .05 level of statistical significance.

Also, according to the OECD, the PISA mathematics framework was revised in 2003; thus, it is not possible to compare mathematics learning outcomes from PISA 2000 with those from PISA 2003, 2006, and 2009. The PISA science framework was revised in 2006; thus, it is not possible to compare science learning outcomes from PISA 2000 and 2003 with those from PISA 2006 and 2009. (NCES Highlights from PISA 2009, pp. 21 and 27)

Endnotes

[1] TIMSS tested 12th grade students in mathematics and science in 1995 and 2008; it also tested additional grades (3, 7, and 8) in 1995. The U.S. participated in the 1995 study, but only 10 countries participated in the 2008 assessment, which was for advanced mathematics and physics. See http://timss.bc.edu for more information about each TIMSS project.

[2] Ina V.S. Mullis, Michael O. Martin, Ann M. Kennedy, and Pierre Foy, PIRLS 2006 International Report, TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College, 2007, p. 15, http://timss.bc.edu (accessed Feb. 16, 2011).

[3] P. Gonzales, T. Williams, L. Jocelyn, S. Roey, D. Kastberg, and S. Brenwald, *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009–001 Revised), National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC, 2008, p. 1, http://nces.ed.gov (accessed Jan. 25, 2011).

[4] Organization for Economic Co-operation and Development, PISA—the OECD Program for International Student Assessment, no date, p. 20, http://www.oecd.org (accessed Jan. 25, 2011).

[5] Organization for Economic Cooperation and Development, *Learning for Tomorrow's World: First Results from PISA 2003*, p. 198, http://www.oecd.org (accessed Feb. 16, 2011).

[6] OECD, *Strong Performers and Successful Reformers in Education: Lessons from PISA for the United States*, 2010, p. 26, http://www.pisa.oecd.org. "Note, however, that because of the way in which the sample was drawn, the performance estimates for the groups of states are associated with considerable error."

[7] P. Gonzales, T. Williams, L. Jocelyn, S. Roey, D. Kastberg, and S. Brenwald, *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009–001 Revised), National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC, 2008, p. 25, http://nces.ed.gov (accessed Jan. 25, 2011).

[8] National Center on Education Statistics, "Trends in International Mathematics and Science Study (TIMSS)," Frequently Asked Questions About the Assessment, http://nces.ed.gov (accessed March 3, 2011).

[9] William H. Schmit, Curtis C. McKnight, and Senta A. Raizen, *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*, Executive Summary, U.S. National Research Center for the Third International Mathematics and Science Study, Michigan State University, 1997, http://ustimss.msu.edu (accessed Jan. 28, 2011).

[10] Center for K-12 Assessment and Performance Management at Educational Testing Service, Coming Together to Raise Achievement: New Assessments for the Common Core State Standards, Dec. 2010, p. 3, http://www.ode.state.or.us (accessed March 1, 2011).

11 Sayuri Takahira, Patrick Gonzales, May Frase, and Laura Hersh Salganik, *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context,* U.S. Department of Education, National Center for Education Statistics, NCES 98-049, Washington, DC, U.S. Government Printing Office, Feb. 1998 (Revised Aug. 1998), p. 66, http://nces.ed.gov (accessed Feb. 2, 2011) .

12 Howard L. Fleischman, Paul J. Hopstock, Marisa P. Pelczar, Brooke E. Shelley, *Highlights from PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context*, U.S. Department of Education, National Center for Educational Statistics, Washington, DC, U.S. Government Printing Office, NCES 2011-004, Dec. 2010, p. 20, http://nces.ed.gov (accessed Feb. 1, 2011).

13 Howard L. Fleischman, Paul J. Hopstock, Marisa P. Pelczar, Brooke E. Shelley, *Highlights from PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context*, U.S. Department of Education, National Center for Educational Statistics, Washington, DC, U.S. Government Printing Office, NCES 2011-004, Dec. 2010, p. 20, http://nces.ed.gov (accessed Feb. 1, 2011).

14 OECD, *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I),* 2010, p. 221, http://browse.oecdbookshop.org/oecd/pdfs/browseit/9810071E.PDF (accessed March 4, 2011).

15 International Association for the Evaluation of Educational Achievement, *TIMSS 2007 Encyclopedia: A Guide to Mathematics and Science Education Around the World* (Volumes 1 and 2), TIMSS and PIRLS International Study Center, Boston College, 2008, p. 4, http://timss.bc.edu (accessed Jan. 26, 2010) and OECD, *Strong Performers and Successful Reformers in Education: Lessons from PISA for the United States*, 2010, p. 22, http://www.oecd.org (accessed Feb. 1, 2011).

16 Mark Schneider, Commissioner, National Center for Education Statistics, Benefits and Limitations of States Benchmarking to International Standards: A Meeting to Assist States in Making Informed Decisions about Participating in International Assessments, Introductory Remarks, http://nces.ed.gov (accessed Feb. 2, 2011).

17 National Center for Education Statistics, National Assessment of Educational Progress, About Us, http://nces.ed.gov (accessed Feb. 3, 2011).

18 National Center for Education Statistics, *Comparing TIMSS with NAEP and PISA in Mathematics and Science*, 2007, http://nces.ed.gov (accessed Feb. 3, 2011).

19 National Center for Education Statistics, *Comparing TIMSS with NAEP and PISA in Mathematics and Science*, no date, p. 5, http://nces.ed.gov (accessed March 8, 2011).

20 Ibid., p. 7.

21 National Center for Education Statistics, Trends in International Mathematics and Science Study International Data Explorer Help Guide, Oct. 2010, p. 3, http://nces.ed.gov (accessed March 1, 2011).

22 Mark Schneider, International Benchmarking, American Institutes for Research, June 2, 2009, p. 15, https://edsurveys.rti.org/PISA (accessed March 1, 2011).

23 As cited by NCES at http://nces.ed.gov/timss/naeplink.asp: E. G. Johnson, Linking the National Assessment of Educational Progress (NAEP) and The Third International Mathematics and Science Study (TIMSS): Eighth-Grade Results (NCES 98-500). National Center for Education Statistics, U.S. Department of Education; Washington, D.C., 1998; E. Johnson, J. Cohen, W. H. Chen, T. Jiang, and Y. Zhang, 2000 NAEP-1999 TIMSS Linking Report (NCES 2005–01), National Center for Education Statistics, U.S. Department of Education. Washington, D.C., 2003; and G. W. Phillips, The Second Derivative: International Benchmarks in Mathematics for American States and School Districts, Washington, D.C., American Institutes for Research, 2009.

24 Eric A. Hanushek, Paul E. Peterson, Ludger Woessmann, *U.S. Math Performance in Global Perspective: How well does each state do at producing high-achieving students?*, Harvard's Program on Education Policy and Governance and Education Next, Taubman Center for State and Local Government, Harvard Kennedy School, PEPG Report No. 10-19, Nov. 2010, p. 11, http://www.hks.harvard.edu (accessed Feb. 6, 2011).

25 National Center for Education Statistics, Trends in International Mathematics and Science Study (TIMSS), "2011 NAEP-TIMSS Linking Study," http://nces.ed.gov (accessed Feb. 7, 2011).

26 National Center for Education Statistics, Trends in International Mathematics and Science Study (TIMSS), "2011 NAEP-TIMSS Linking Study," http://nces.ed.gov (accessed Feb. 7, 2011).

Participants in PIRLS, 2001 and 2006

Countries	2006	2001
Argentina		●
Austria	●	
Belgium (Flemish)	●	
Belgium (French)	●	
Belize		●
Bulgaria	●	●
Canada, Alberta	●	
Canada, British Columbia	●	
Canada, Nova Scotia	●	
Canada, Ontario	●	●
Canada, Quebec	●	●
Chinese Taipei	●	
Colombia		●
Cyprus		●
Czech Republic		●
Denmark	●	
England	●	●
France	●	●
Georgia	●	
Germany	●	●
Greece		●
Hong Kong SAR	●	●
Hungary	●	●
Iceland	●	●
Indonesia	●	
Iran, Islamic Rep. of	●	●
Israel	●	●
Italy	●	●
[1] Kuwait	●	
Latvia	●	●
Lithuania	●	●
Luxembourg	●	
Macedonia, Rep. of	●	●
Moldova, Rep. of	●	●
Morocco	●	●
Netherlands	●	●
New Zealand	●	●
Norway	●	●
Poland	●	
Qatar	●	
Romania	●	●
Russian Federation	●	●
Scotland	●	●
Singapore	●	●
Slovak Republic	●	●
Slovenia	●	●
South Africa	●	
Spain	●	
Sweden	●	●
Trinidad and Tobago	●	
Turkey		●
United States	●	●

Indicates country participation in that testing cycle ●

1 Although Kuwait participated in PIRLS 2001, the data were not considered comparable for measuring trends.

Source: International Association for the Evaluation of Educational Achievement, *PIRLS 2006 Technical Report*, Edited by Michael O. Martin, Ina V.S Mullis, and Ann M. Kennedy, TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College, 2007, p. 3, http://timss.bc.edu (accessed Feb. 18, 2011).

Participants in TIMSS 1995 through 2007

Country	Grade 4			Grade 8			
	2007	2003	1995	2007	2003	1999	1995
Algeria	●			●			
Armenia	●	●		●	●		
Australia	●	●	●	●	●	●	●
Austria	●		●				●
Bahrain				●	●		
Bosnia and Herzegovina				●			
Botswana				●	●		
Bulgaria				●	●	●	●
Chinese Taipei	●	●		●	●	●	
Colombia	●			●			●
Cyprus		●	●	●	●	●	●
Czech Republic	●		●	●		●	●
Denmark	●						●
Egypt				●	●		
El Salvador	●			●			
England	●	●	●	●	●	●	●
Georgia	●			●			
Germany	●						●
Ghana				●	●		
Hong Kong SAR	●	●	●	●	●	●	●
Hungary	●	●	●	●	●	●	●
Indonesia				●	●	●	
Iran, Islamic Rep. of	●	●	●	●	●	●	●
Israel			●	●	●	●	●
Italy	●	●	●	●	●	●	●
Japan	●	●	●	●	●	●	●
Jordan				●	●	●	
Kazakhstan	●						
Korea, Rep. of			●	●	●	●	●
Kuwait	●		●	●			●
Latvia	●	●	●		●	●	●
Lebanon				●	●		
Lithuania	●	●		●	●	●	●
Malaysia				●	●	●	
Malta				●			
Mongolia	●			●			
Morocco	●	●		●	●	●	
Netherlands	●	●	●		●	●	●
New Zealand	●	●	●		●	●	●
Norway	●	●	●	●	●		●
Oman				●			
Palestinian Nat'l Auth.				●	●		
Qatar	●			●			
Romania				●	●	●	●
Russian Federation	●	●		●	●	●	●
Saudi Arabia				●	●		
Scotland	●	●	●	●	●		●
Serbia				●	●		
Singapore	●	●	●	●	●	●	●
Slovak Republic	●				●	●	●
Slovenia	●	●	●	●	●	●	●
Sweden	●			●	●		●
Syrian Arab Republic				●	●		
Thailand			●	●		●	●
Tunisia	●	●		●	●	●	
Turkey				●		●	
Ukraine	●			●			
United States	●	●	●	●	●	●	●
Yemen	●	●					
Benchmarking Participants							
Alberta, Canada	●		●			●	●
Basque Country, Spain				●	●		
British Columbia, Canada	●			●		●	
Dubai, UAE	●			●			
Massachusetts, US	●			●		●	
Minnesota, US	●		●	●			●
Ontario, Canada	●	●	●	●	●	●	●
Quebec, Canada	●	●	●	●	●	●	●

SOURCE: IEA's Trends in International Mathematics and Science Study (TIMSS) 2007

Source: Ina V.S. Mullis, Michael O. Martin, and Pierre Foy, *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*, TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College, 2008, pp. 20-21, http://timss.bc.edu (accessed Feb. 18, 2011).

Participants in PISA, 2000 through 2009

Country	2000	2003	2006	2009
OECD countries				
Australia	•	•	•	•
Austria	•	•	•	•
Belgium	•	•	•	•
Canada	•	•	•	•
Chile	•		•	•
Czech Republic	•	•	•	•
Denmark	•	•	•	•
Estonia			•	•
Finland	•	•	•	•
France	•	•	•	•
Germany	•	•	•	•
Greece	•	•	•	•
Hungary	•	•	•	•
Iceland	•	•	•	•
Ireland	•	•	•	•
Israel	•		•	•
Italy	•	•	•	•
Japan	•	•	•	•
Korea, Republic of	•	•	•	•
Luxembourg	•	•	•	•
Mexico	•	•	•	•
Netherlands	•	•	•	•
New Zealand	•	•	•	•
Norway	•	•	•	•
Poland	•	•	•	•
Portugal	•	•	•	•
Slovak Republic		•	•	•
Slovenia			•	•
Spain	•	•	•	•
Sweden	•	•	•	•
Switzerland	•	•	•	•
Turkey		•	•	•
United Kingdom	•	•	•	•
United States	•	•	•	•

Country	2000	2003	2006	2009
Non-OECD countries				
Albania	•			•
Argentina	•		•	•
Azerbaijan			•	•
Brazil	•	•	•	•
Bulgaria	•		•	•
Chinese Taipei			•	•
Colombia			•	•
Croatia			•	•
Dubai-UAE				•
Hong Kong-China	•	•	•	•
Indonesia	•	•	•	•
Jordan			•	•
Kazakhstan				•
Kyrgyz Republic			•	•
Latvia	•	•	•	•
Liechtenstein	•	•	•	•
Lithuania			•	•
Macao-China		•	•	•
Macedonia	•			
Montenegro, Republic of [1]		•	•	•
Panama				•
Peru	•			•
Qatar			•	•
Romania	•		•	•
Russian Federation	•	•	•	•
Serbia, Republic of [1]		•	•	•
Shanghai-China				•
Singapore				•
Thailand	•	•	•	•
Trinidad and Tobago				•
Tunisia		•	•	•
Uruguay		•	•	•

[1] The Republics of Montenegro and Serbia were a united jurisdiction under the PISA 2003 assessment.

NOTE: A "•" indicates that the country participated in the Program for International Student Assessment (PISA) in the specific year. Because PISA is principally an Organization for Economic Cooperation and Development (OECD) study, non-OECD countries are displayed separately from the OECD countries. Eleven countries and other education systems—Albania, Argentina, Bulgaria, Chile, Hong Kong-China, Indonesia, Israel, Macedonia, Peru, Romania, and Thailand—administered PISA 2000 in 2001. Italics indicate non-national entities. UAE refers to the United Arab Emirates.

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000, 2003, 2006, and 2009.

Source: Howard L. Fleischman, Paul J. Hopstock, Marisa P. Pelczar, Brooke E. Shelley, *Highlights from PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context*, U.S. Department of Education, National Center for Educational Statistics, Washington, DC, U.S. Government Printing Office, NCES 2011-004, Dec. 2010, p. 3, http://nces.ed.gov (accessed Feb. 18, 2011).

Description of PIRLS international benchmarks: 2006

Benchmark	Cutpoint	Reading skills and strategies
Advanced	625	▪ Interpret figurative language ▪ Distinguish and interpret complex information from different parts of text ▪ Integrate ideas across text to provide interpretations about character's feelings and behaviors
High	550	▪ Recognize some textual features, such as figurative language and abstract messages ▪ Make inferences on the basis of abstract or embedded information ▪ Integrate information to recognize main ideas and provide explanations
Intermediate	475	▪ Identify central events, plot sequences, and relevant story details ▪ Make straightforward inferences from the text ▪ Begin to make connections across parts of the text
Low	400	▪ Retrieve explicitly stated details from literary and informational texts

Source: J. Baer, S. Baldi, K. Ayotte, and P. Green, *The Reading Literacy of U.S. Fourth-Grade Students in an International Context: Results From the 2001 and 2006 Progress in International Reading Literacy Study (PIRLS),* NCES 2008–017, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C., 2007, p. 4, http://nces.ed.gov (accessed Feb. 8, 2011).

Description of TIMSS international mathematics benchmarks, by grade: 2007

Benchmark / score cutpoint	Grade 4
Advanced / 625	*Students can apply their understanding and knowledge in a variety of relatively complex situations and explain their reasoning.* They can apply proportional reasoning in a variety of contexts. They demonstrate a developing understanding of fractions and decimals. They can select appropriate information to solve multistep word problems. They can formulate or select a rule for a relationship. Students can apply geometric knowledge of a range of two- and three-dimensional shapes in a variety of situations. They can organize, interpret, and represent data to solve problems.
High / 550	*Students can apply their knowledge and understanding to solve problems.* Students can solve multistep word problems involving operations with whole numbers. They can use division in a variety of problem situations. They demonstrate understanding of place value and simple fractions. Students can extend patterns to find a later specified term and identify the relationship between ordered pairs. Students show some basic geometric knowledge. They can interpret and use data in tables and graphs to solve problems.
Intermediate / 475	*Students can apply basic mathematical knowledge in straightforward situations.* Students at this level demonstrate an understanding of whole numbers. They can extend simple numeric and geometric patterns. They are familiar with a range of two-dimensional shapes. They can read and interpret different representations of the same data.
Low / 400	*Students have some basic mathematical knowledge.* Students can demonstrate an understanding of adding and subtracting with whole numbers. They demonstrate familiarity with triangles and informal coordinate systems. They can read information from simple bar graphs and tables

Benchmark / score cutpoint	Grade 8
Advanced / 625	*Students can organize and draw conclusions from information, make generalizations, and solve nonroutine problems.* They can solve a variety of ratio, proportion, and percent problems. They can apply their knowledge of numeric and algebraic concepts and relationships. Students can express generalizations algebraically and model situations. They can apply their knowledge of geometry in complex problem situations. Students can derive and use data from several sources to solve multistep problems.
High / 550	*Students can apply their understanding and knowledge in a variety of relatively complex situations.* They can relate and compute with fractions, decimals, and percents, operate with negative integers, and solve word problems involving proportions. Students can work with algebraic expressions and linear equations. Students use knowledge of geometric properties to solve problems, including area, volume, and angles. They can interpret data in a variety of graphs and table and solve simple problems involving probability.
Intermediate / 475	*Students can apply basic mathematical knowledge in straightforward situations.* They can add and multiply to solve one-step word problems involving whole numbers and decimals. They can work with familiar fractions. They understand simple algebraic relationships. They demonstrate understanding of properties of triangles and basic geometric concepts. They can read and interpret graphs and tables. They recognize basic notions of likelihood.
Low / 400	*Students have some knowledge of whole numbers and decimals, operations, and basic graphs.*

Note: Score cutpoints for the international benchmarks are determined through scale anchoring. Scale anchoring involves selecting benchmarks (scale points) on the achievement scales to be described in terms of student performance, and then identifying items that students scoring at the anchor points can answer correctly. The score cutpoints are set at equal intervals along the achievement scales. The score cutpoints were selected to be as close as possible to the standard percentile cutpoints (i.e., 90th, 75th, 50th, and 25th percentiles).
Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Description of TIMSS international science benchmarks, by grade: 2007

Benchmark / score cutpoint	Grade 4
Advanced / 625	*Students can apply knowledge and understanding of scientific processes and relationships in beginning scientific inquiry.* Students communicate their understanding of characteristics and life processes of organisms as well as of factors relating to human health. They demonstrate understanding of relationships among various physical properties of common materials and have some practical knowledge of electricity. Students demonstrate some understanding of the solar system and Earth's physical features and processes. They show a developing ability to interpret the results of investigations and draw conclusions as well as a beginning ability to evaluate and support an argument.
High / 550	*Students can apply knowledge and understanding to explain everyday phenomena.* Students demonstrate some understanding of plant and animal structure, life processes, and the environment and some knowledge of properties of matter and physical phenomena. They show some knowledge of the solar system, and of Earth's structure, processes, and resources. Students demonstrate beginning scientific inquiry knowledge and skills, and provide brief descriptive responses combining knowledge of science concepts with information from everyday experience of physical and life processes.
Intermediate / 475	*Students can apply basic knowledge and understanding to practical situations in the sciences.* Students recognize some basic information related to characteristics of living things and their interaction with the environment, and show some understanding of human biology and health. They also show some understanding of familiar physical phenomena. Students know some basic facts about the solar system and have a developing understanding of Earth's resources. They demonstrate some ability to interpret information in pictorial diagrams and apply factual knowledge to practical situations.
Low / 400	*Students have some elementary knowledge of life science and physical science.* Students can demonstrate knowledge of some simple facts related to human health and the behavioral and physical characteristics of animals. They recognize some properties of matter, and demonstrate a beginning understanding of forces. Students interpret labeled pictures and simple diagrams, complete simple tables, and provide short written responses to questions requiring factual information.

Benchmark / score cutpoint	Grade 8
Advanced / 625	*Students can demonstrate a grasp of some complex and abstract concepts in biology, chemistry, physics, and Earth science.* They have an understanding of the complexity of living organisms and how they relate to their environment. They show understanding of the properties of magnets, sound, and light, as well as demonstrating understanding the structure of matter and physical and chemical properties and changes. Students apply knowledge of the solar system and of Earth's features and processes, and apply understanding of major environmental issues. They understand some fundamentals of scientific investigation and can apply basic physical principles to solve some quantitative problems. They can provide written explanations to communicate scientific knowledge.
High / 550	*Students can demonstrate conceptual understanding of some science cycles, systems, and principles.* They have some understanding of biological concepts including cell processes, human biology and health, and the interrelationship of plants and animals in ecosystems. They apply knowledge to situations related to light and sound, demonstrate elementary knowledge of heat and forces, and show some evidence of understanding the structure of matter, and chemical and physical properties and changes. They demonstrate some understanding of the solar system, Earth's processes and resources, and some basic understanding of major environmental issues. Students demonstrate some scientific inquiry skills. They combine information to draw conclusions,interpret tabular and graphical information, and provide short explanations conveying scientific knowledge.
Intermediate / 475	*Students can recognize and communicate basic scientific knowledge across a range of topics.* They demonstrate some understanding of characteristics of animals, food webs, and the effect of population changes in ecosystems. They are acquainted with some aspects of sound and force and have elementary knowledge of chemical change. They demonstrate elementary knowledge of the solar system, Earth's processes, and resources and the environment. Students extract information from tables and interpret pictorial diagrams. They can apply knowledge to practical situations and communicate their knowledge through brief descriptive responses.
Low / 400	*Students can recognize some basic facts from the life and physical sciences.* They have some knowledge of the human body, and demonstrate some familiarity with everyday physical phenomena. Students can interpret pictorial diagrams and apply knowledge of simple physical concepts to practical situations.

Note: Score cutpoints for the international benchmarks are determined through scale anchoring. Scale anchoring involves selecting benchmarks (scale points) on the achievement scales to be described in terms of student performance, and then identifying items that students scoring at the anchor points can answer correctly. The score cutpoints are set at equal intervals along the achievement scales. The score cutpoints were selected to be as close as possible to the standard percentile cutpoints (i.e., 90th, 75th, 50th, and 25th percentiles).
Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Description of PISA proficiency levels on combined reading literacy scale: 2009

Proficiency level and lower cut point score	Task descriptions
Level 6 686	At level 6, tasks typically require the reader to make multiple inferences, comparisons and contrasts that are both detailed and precise. They require demonstration of a full and detailed understanding of one or more texts and may involve integrating information from more than one text. Tasks may require the reader to deal with unfamiliar ideas, in the presence of prominent competing information, and to generate abstract categories for interpretations. Reflect and evaluate tasks may require the reader to hypothesize about or critically evaluate a complex text on an unfamiliar topic, taking into account multiple criteria or perspectives, and applying sophisticated understandings from beyond the text. There are limited data about access and retrieve tasks at this level, but it appears that a salient condition is precision of analysis and fine attention to detail that is inconspicuous in the texts.
Level 5 626	At level 5, tasks involve retrieving information that require the reader to locate and organize several pieces of deeply embedded information, inferring which information in the text is relevant. Reflective tasks require critical evaluation or hypothesis, drawing on specialized knowledge. Both interpretative and reflective tasks require a full and detailed understanding of a text whose content or form is unfamiliar. For all aspects of reading, tasks at this level typically involve dealing with concepts that are contrary to expectations.
Level 4 553	At level 4, tasks involve retrieving information that require the reader to locate and organize several pieces of embedded information. Some tasks at this level require interpreting the meaning of nuances of language in a section of text by taking into account the text as a whole. Other interpretative tasks require understanding and applying categories in an unfamiliar context. Reflective tasks at this level require readers to use formal or public knowledge to hypothesize about or critically evaluate a text. Readers must demonstrate an accurate understanding of long or complex texts whose content or form may be unfamiliar.
Level 3 480	At level 3, tasks require the reader to locate, and in some cases recognize the relationship between, several pieces of information that must meet multiple conditions. Interpretative tasks at this level require the reader to integrate several parts of a text in order to identify a main idea, understand a relationship or construe the meaning of a word or phrase. They need to take into account many features in comparing, contrasting or categorizing. Often the required information is not prominent or there is much competing information; or there are other text obstacles, such as ideas that are contrary to expectation or negatively worded. Reflective tasks at this level may require connections, comparisons, and explanations, or they may require the reader to evaluate a feature of the text. Some reflective tasks require readers to demonstrate a fine understanding of the text in relation to familiar, everyday knowledge. Other tasks do not require detailed text comprehension but require the reader to draw on less common knowledge.
Level 2 407	At level 2, some tasks require the reader to locate one or more pieces of information, which may need to be inferred and may need to meet several conditions. Others require recognizing the main idea in a text, understanding relationships, or construing meaning within a limited part of the text when the information is not prominent and the reader must make low level inferences. Tasks at this level may involve comparisons or contrasts based on a single feature in the text. Typical reflective tasks at this level require readers to make a comparison or several connections between the text and outside knowledge, by drawing on personal experience and attitudes.

Level 1a 335	At level 1a, tasks require the reader to locate one or more independent pieces of explicitly stated information; to recognize the main theme or author's purpose in a text about a familiar topic, or to make a simple connection between information in the text and common, everyday knowledge. Typically the required information in the text is prominent and there is little, if any, competing information. The reader is explicitly directed to consider relevant factors in the task and in the text.
Level 1b 262	At level 1b, tasks require the reader to locate a single piece of explicitly stated information in a prominent position in a short, syntactically simple text with a familiar context and text type, such as a narrative or a simple list. The text typically provides support to the reader, such as repetition of information, pictures or familiar symbols. There is minimal competing information. In tasks requiring interpretation the reader may need to make simple connections between adjacent pieces of information.

Note: To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into reading literacy levels according to their scores. Cut point scores in the exhibit are rounded; exact cut point scores are provided in appendix B.

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

Description of PISA proficiency levels on mathematics literacy scale: 2009

Proficiency level and lower cut point score	Task descriptions
Level 6 669	At level 6, students can conceptualize, generalize, and utilize information based on their investigations and modeling of complex problem situations. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply this insight and understandings along with a mastery of symbolic and formal mathematical operations and relationships to develop new approaches and strategies for attacking novel situations. Students at this level can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments, and the appropriateness of these to the original situations.
Level 5 607	At level 5, students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare, and evaluate appropriate problem solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterizations, and insight pertaining to these situations. They can reflect on their actions and formulate and communicate their interpretations and reasoning.
Level 4 545	At level 4, students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic ones, linking them directly to aspects of real-world situations. Students at this level can utilize well-developed skills and reason flexibly, with some insight, in these contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments, and actions.
Level 3 482	At level 3, students can execute clearly described procedures, including those that require sequential decisions. They can select and apply simple problem solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They can develop short communications reporting their interpretations, results and reasoning.
Level 2 420	At level 2, students can interpret and recognize situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures, or conventions. They are capable of direct reasoning and making literal interpretations of the results.

Level 1 358	At level 1, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli.

Note: To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into mathematics literacy levels according to their scores. Cut point scores in the exhibit are rounded; exact cut point scores are provided in appendix B. SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

Description of PISA proficiency levels on science literacy scale: 2009

Proficiency level and lower cut point score	Task descriptions
Level 6 708	At level 6, students can consistently identify, explain and apply scientific knowledge and knowledge about science in a variety of complex life situations. They can link different information sources and explanations and use evidence from those sources to justify decisions. They clearly and consistently demonstrate advanced scientific thinking and reasoning, and they demonstrate willingness to use their scientific understanding in support of solutions to unfamiliar scientific and technological situations. Students at this level can use scientific knowledge and develop arguments in support of recommendations and decisions that center on personal, social or global situations.
Level 5 633	At level 5, students can identify the scientific components of many complex life situations, apply both scientific concepts and knowledge about science to these situations, and can compare, select and evaluate appropriate scientific evidence for responding to life situations. Students at this level can use well-developed inquiry abilities, link knowledge appropriately and bring critical insights to situations. They can construct explanations based on evidence and arguments based on their critical analysis.
Level 4 559	At level 4, students can work effectively with situations and issues that may involve explicit phenomena requiring them to make inferences about the role of science or technology. They can select and integrate explanations from different disciplines of science or technology and link those explanations directly to aspects of life situations. Students at this level can reflect on their actions and they can communicate decisions using scientific knowledge and evidence.
Level 3 484	At level 3, students can identify clearly described scientific issues in a range of contexts. They can select facts and knowledge to explain phenomena and apply simple models or inquiry strategies. Students at this level can interpret and use scientific concepts from different disciplines and can apply them directly. They can develop short statements using facts and make decisions based on scientific knowledge.
Level 2 410	At level 2, students have adequate scientific knowledge to provide possible explanations in familiar contexts or draw conclusions based on simple investigations. They are capable of direct reasoning and making literal interpretations of the results of scientific inquiry or technological problem solving.
Level 1 335	At level 1, students have such a limited scientific knowledge that it can only be applied to a few, familiar situations. They can present scientific explanations that are obvious and follow explicitly from given evidence.

Note: To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into mathematics literacy levels according to their scores. Cut point scores in the exhibit are rounded; exact cut point scores are provided in appendix B.
Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

International testing organizations

OECD Program for International Student Assessment (PISA)
http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html

Select publications
Strong Performers and Successful Reformers: Lessons from PISA for the United States
http://www.pisa.oecd.org/dataoecd/32/50/46623978.pdf

PISA 2009 Results—Five volumes:
http://www.pisa.oecd.org/document/61/0,3746,en_32252351_32235731_46567613_1_1_1_1,00.html
Vol I. *What Students Know and Can Do: Student Performance in Reading, Mathematics, and Science*
Vol. II. *Overcoming Social Background: Equity in Learning Opportunities and Outcomes*
Vol. III. *Learning to Learn—Student Engagement, Strategies and Practices*
Vol. IV. *What Makes a School Successful: Resources, Policies and Practices*
Vol. V. *Learning Trends: Changes in Student Performance Since 2000*

IEA Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) http://timssandpirls.bc.edu

Select publications
All reports available at http://timssandpirls.bc.edu/isc/publications.html
TIMSS 2007 International Mathematics Report
TIMSS 2007 International Science Report
TIMSS 2007 Encyclopedia
TIMSS 2007 Assessment Frameworks
PIRLS 2006 International Report
PIRLS 2006 Encyclopedia
PIRLS 2006 Technical Report
PIRLS 2006 Assessment Framework and Specifications

National Center for Education Statistics (NCES)

NCES page for each of the three international assessments:
Progress in International Reading Literacy Study (PIRLS) http://nces.ed.gov/surveys/pirls
Trends in International Mathematics and Science Study (TIMSS) http://nces.ed.gov/timss
Program for International Student Assessment (PISA) http://nces.ed.gov/surveys/pisa

Select publications
The Reading Literacy of U.S. Fourth-Grade Students in an International Context: Results From the 2001 and 2006 Progress in International Reading Literacy Study, 2008 (NCES 2008-017)
http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008017 (NCES web page)
http://nces.ed.gov/pubs2008/2008017.pdf (document)

U.S. Performance Across International Assessments of Student Achievement: Special Supplement to The Condition of Education 2009, 2009 (NCES 2009-083)
http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009083 (NCES web page)
http://nces.ed.gov/pubs2009/2009083.pdf (document)

Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context, Sept. 2009 (NCES 2009-001)
http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001 (NCES web page)
http://nces.ed.gov/pubs2009/2009001.pdf (document)

Highlights from PISA 2009: Performance of U.S. 15-Year-Olds in Reading, Mathematics, and Science Literacy in an International Context, Dec. 2010 (NCES 2011-004)
http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011004 (NCES web page)
http://nces.ed.gov/pubs2011/2011004.pdf (document)

Other items of interest

Linking international test results to NAEP

Gary W. Phillips, Ph.D., Vice President and Chief Scientist, American Institute for Research, *The Second Derivative: International Benchmarks in Mathematics for U.S. States and School Districts*, June 2009
http://www.air.org/expertise/index/?fa=viewContent&content_id=29 (web page with links to related videos)
http://www.air.org/files/International_Benchmarks1.pdf (document)

Eric A. Hanushek, Paul E. Peterson, Ludger Woessmann, *U.S. Math Performance in Global Perspective: How well does each state do at producing high-achieving students?*, Harvard's Program on Education Policy and Governance and Education Next, Taubman Center for State and Local Government, Harvard Kennedy School, PEPG Report No. 10-19, Nov. 2010,
http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG10-19_HanushekPetersonWoessmann.pdf

National Center for Education Statistics, "2010-11 NAEP-TIMSS Linking Study"
http://nces.ed.gov/timss/naeplink.asp

Criticism of international tests

Tom Loveless, Senior Fellow, Brown Center on Education Policy, *The 2008 Brown Center Report on American Education: How Well Are American Students Learning?*, Jan. 2009, Vol. II, No. 3,
http://www.brookings.edu/~/media/Files/rc/reports/2009/0225_education_loveless/0225_education_loveless.pdf

B. Lindsay Lowell and Hal Salzman, *Into the Eye of the Storm: Assessing the Evidence on Science and Engineering Education, Quality, and Workforce Demand*, The Urban Institute, Oct. 2007,
http://www.urban.org/UploadedPDF/411562_Salzman_Science.pdf

Mark Schneider, *International Benchmarking*, American Institutes for Research, June 2, 2009,
https://edsurveys.rti.org/PISA/documents/schneiderNCES_International_Benchmarking_final.pdf

Ben Wildavsky, "Think Again: Education," *Foreign Policy*, March/April 2011,
http://www.foreignpolicy.com/articles/2011/02/22/think_again_education?page=full

Offices of Research and Education Accountability
Phillip Doss, Director
Suite 1700, James K. Polk Building ▪ 505 Deaderick Street
Nashville, Tennessee 37243 ▪ (615) 401-7911
www.tn.gov/comptroller/orea