NCES 2011-004
U.S. DEPARTMENT OF EDUCATION

Highlights From PISA 2009:

Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

Highlights From PISA 2009:

Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context

December 2010

Howard L. Fleischman
Paul J. Hopstock
Marisa P. Pelczar
Brooke E. Shelley
Windwalker Corporation

Holly Xie
Project Officer
National Center for Education Statistics

NCES 2011-004
U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Statistics
Stuart Kerachsky
Acting Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

NCES, IES, U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

December 2010

The NCES Home Page address is http://nces.ed.gov.
The NCES Publications and Products address is http://nces.ed.gov/pubsearch.

This report was prepared for the National Center for Education Statistics under Contract No. ED-04-CO-0084 with Windwalker Corporation. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Suggested Citation
Fleischman, H.L., Hopstock, P.J., Pelczar, M.P., and Shelley, B.E. (2010). *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. Washington, DC: U.S. Government Printing Office.

For ordering information on this report, write to
ED Pubs, U.S. Department of Education
P.O. Box 22207
Alexandria, VA 22304

Or call toll free 1-877-4-ED-Pubs or order online at http://www.edpubs.gov.

Content Contact
Holly Xie
(202) 502-7314
holly.xie@ed.gov

Executive Summary

The Program for International Student Assessment (PISA) is an international assessment that measures the performance of 15-year-olds in reading literacy, mathematics literacy, and science literacy every 3 years. First implemented in 2000, PISA is coordinated by the Organization for Economic Cooperation and Development (OECD), an intergovernmental organization of 34 member countries. In all, 60 countries and 5 other education systems[1] participated as partners in PISA 2009.

Each PISA cycle assesses one of the three subject areas in depth. In PISA 2009, reading literacy was the subject area assessed in depth, and science literacy and mathematics literacy were the minor subjects assessed. This report focuses on the performance of U.S. students[2] in the major subject area of reading literacy by presenting results from a combined reading literacy scale and three reading literacy subscales: *access and retrieve, integrate and interpret,* and *reflect and evaluate*. Achievement results for the minor subject areas of mathematics and science literacy are also presented.

Key findings from PISA 2009 include the following:

Reading Literacy

- U.S. 15-year-olds had an average score of 500 on the combined reading literacy scale, not measurably different from the OECD average score of 493. Among the 33 other OECD countries, 6 countries had higher average scores than the United States, 13 had lower average scores, and 14 had average scores not measurably different from the U.S. average. Among the 64 other OECD countries, non-OECD countries, and other education systems, 9 had higher average scores than the United States, 39 had lower average scores, and 16 had average scores not measurably different from the U.S. average.
- On the *reflect and evaluate* reading literacy subscale, U.S. 15-year-olds had a higher average score than the OECD average. The U.S. average was lower than that of 5 OECD countries and higher than that of 23 OECD countries; it was lower than that of 8 countries and other education systems and higher than that of 51 countries and other education systems overall. On the other two subscales—*access and retrieve* and *integrate and interpret*—the U.S. average was not measurably different from the OECD average.
- In reading literacy, 30 percent of U.S. students scored at or above proficiency level 4. Level 4 is the level at which students are "capable of difficult reading tasks, such as locating embedded information, construing meaning from nuances of language and critically evaluating a text" (OECD 2010a, p. 51). At levels 5 and 6 students demonstrate higher-level reading skills and may be referred to as "top performers" in reading. There was no measurable difference between the percentage of U.S. students and the percentage of students in the OECD countries on average who performed at or above level 4.
- Eighteen percent of U.S. students scored below level 2 in reading literacy. Students performing below level 2 in reading literacy are below what OECD calls "a baseline level of proficiency, at which students begin to demonstrate the reading literacy competencies that will enable them to participate effectively and productively in life" (OECD 2010a, p. 52). There was no measurable difference between the percentage of U.S. students and the percentage of students in the OECD countries on average who demonstrated proficiency below level 2.
- Female students scored higher, on average, than male students on the combined reading literacy scale in all 65 participating countries and other education systems. In the United States, the difference was smaller than the difference in the OECD countries, on average, and smaller than the differences in 45 countries and other education systems (24 OECD countries and 21 non-OECD countries and other education systems).
- On the combined reading literacy scale, White (non-Hispanic) and Asian (non-Hispanic) students had higher average scores than the overall OECD and U.S. average scores, while Black (non-Hispanic) and Hispanic students had lower average scores than the overall OECD and U.S. average scores. The average scores of students who reported two or more races were not measurably different from the overall OECD or U.S. average scores.
- Students in public schools in which half or more of students (50 to 74.9 percent and 75 percent or more) were eligible for free or reduced-price lunch (FRPL-eligible) scored, on average, below the overall OECD and U.S. average scores in reading literacy. Students in

[1] Other education systems are located in non-national entities, such as Shanghai-China.
[2] In the United States, a total of 165 schools and 5,233 students participated in the assessment. The overall weighted school response rate was 68 percent before the use of replacement schools. The final weighted student response rate after replacement was 87 percent.

schools in which less than 25 percent of students were FRPL-eligible (10 to 24.9 percent and less than 10 percent) scored, on average, above the overall OECD and U.S. average scores. The average scores of students in schools in which 25 to 49.9 percent were FRPL-eligible were above the overall OECD average but not measurably different from the U.S. average.

- 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. There also were no measurable differences between the U.S. average score and the OECD average score in 2000 or in 2009.[3]

Mathematics Literacy

- U.S. 15-year-olds had an average score of 487 on the mathematics literacy scale, which was lower than the OECD average score of 496. 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. Among the 64 other OECD countries, non-OECD countries, and other education systems, 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.

- In mathematics literacy, 27 percent of U.S. students scored at or above proficiency level 4. This is lower than the 32 percent of students in the OECD countries on average that scored at or above level 4. Level 4 is the level at which students can complete higher order tasks such as "solv[ing] problems that involve visual and spatial reasoning...in unfamiliar contexts" and "carry[ing] out sequential processes" (OECD 2004, p. 55). Twenty-three percent of U.S. students scored below level 2. There was no measurable difference between the percentage of U.S. students and the percentage of students in the OECD countries on average demonstrating proficiency below level 2, what OECD calls a "a baseline level of mathematics proficiency on the PISA scale at which students begin to demonstrate the kind of literacy skills that enable them to actively use mathematics" (OECD 2004, p. 56).

- 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. U.S. students' average scores were lower than the OECD average scores in each of these years.[4]

Science Literacy

- 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. Among the 64 other OECD countries, non-OECD countries, and other education systems, 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.

- 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. Level 4 is the level at which students can complete higher order tasks such as "select[ing] and integrat[ing] explanations from different disciplines of science or technology and link[ing] those explanations directly to...life situations" (OECD 2007, p. 43). Eighteen percent of U.S. students and students in the OECD countries on average scored below level 2. Students performing below level 2 are below what OECD calls a "baseline level of proficiency…at which students begin to demonstrate the science competencies that will enable them to participate effectively and productively in life situations related to science and technology" (OECD 2007, p. 44). There were no measurable differences between the percentages of U.S. students and students in the OECD countries on average that scored at the individual proficiency levels.

- The U.S. average score in science literacy in 2009

[3] The OECD averages against which the U.S. averages are compared are the averages for the 27 OECD countries with comparable data for 2000 and 2009.

[4] The OECD averages against which the U.S. averages are compared are the averages for the 29 OECD countries with comparable data for 2003 and 2009.

was higher than the U.S. average in 2006, the only time point to which PISA 2009 performance can be compared in science literacy. While U.S. students scored lower than 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.[5]

[5] The OECD averages against which the U.S. averages are compared are the averages for the 34 OECD countries.

Page intentionally left blank

Acknowledgments

This report reflects the contributions of many individuals. The authors wish to thank all those who assisted with PISA 2009, from the design stage through the creation of this report. The members of the U.S. PISA 2009 Steering Committee (noted in appendix C) gave their time and expertise toward reviewing the project. Finally, the authors wish to thank the many principals, school staff members, and students who generously gave their time to participate in PISA 2009.

Page intentionally left blank

Contents

List of Tables

Table **Page**

List of Figures

List of Exhibits

Page intentionally left blank

Page intentionally left blank

Introduction

PISA in Brief

The Program for International Student Assessment (PISA) is an international assessment that measures the performance of 15-year-olds in reading literacy, mathematics literacy, and science literacy. Coordinated by the Organization for Economic Cooperation and Development (OECD), an intergovernmental organization of 34 member countries, PISA was first implemented in 2000 and is conducted every 3 years. PISA 2009 was the fourth cycle of the assessment.

Each PISA data collection effort assesses one of the three subject areas in depth (considered the major subject area), although all three are assessed in each cycle (the other two subjects are considered minor subject areas for that assessment year). Assessing all three areas allows participating countries to have an ongoing source of achievement data in every subject area while rotating one area as the main focus over the years. In the fourth cycle of PISA, reading was the subject area assessed in depth, as it was in 2000 (figure 1).

Sixty countries and 5 other education systems[1] participated as partners in PISA 2009 (figure 2 and table 1).

This report focuses on the performance of U.S. students in the major subject area of reading literacy as assessed in PISA 2009. Achievement results for the minor subject areas of mathematics and science literacy in 2009 are also presented.

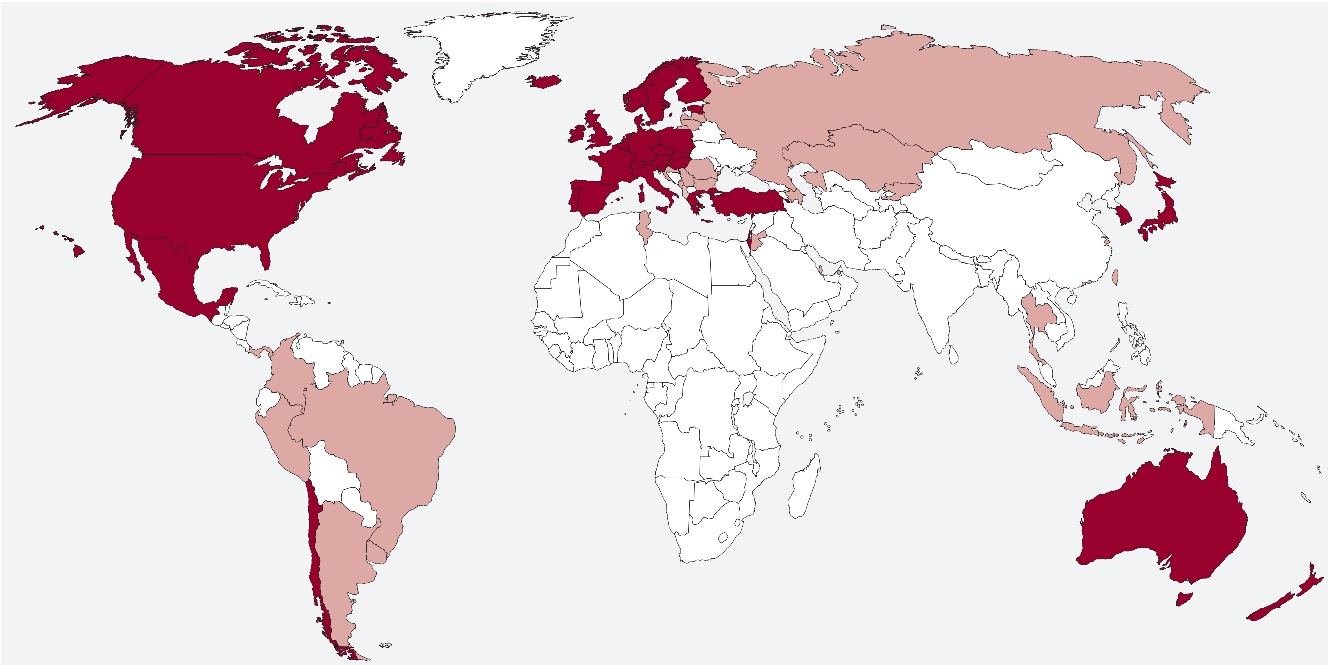
[1] Other education systems are located in non-national entities, such as Shanghai-China.

Figure 1. PISA administration cycle

Assessment year	2000	2003	2006	2009	2012	2015
Subjects assessed	**READING** Mathematics Science	Reading **MATHEMATICS** Science Problem solving	Reading Mathematics **SCIENCE**	**READING** Mathematics Science	Reading **MATHEMATICS** Science Problem solving	Reading Mathematics **SCIENCE**

NOTE: Reading, mathematics, and science literacy are all assessed in each assessment cycle of the Program for International Student Assessment (PISA). A separate problem-solving assessment was administered in 2003 and is planned for 2012. The subject in all capital letters is the major subject area for that cycle.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

Figure 2. Countries that participated in PISA 2009



■ OECD country ■ Non-OECD country or non-national entity □ Non-participating country

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

Table 1. Participation in PISA, by country: 2000, 2003, 2006, and 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.

What PISA Measures

PISA assesses the application of knowledge in reading, mathematics, and science literacy to problems within a real-life context (OECD 1999). PISA uses the term "literacy" in each subject area to denote its broad focus on the application of knowledge and skills. For example, when assessing reading, PISA assesses how well 15-year-old students can understand, use, and reflect on written text for a variety of purposes and settings. In science, PISA assesses how well students can apply scientific knowledge and skills to a range of different situations they may encounter in their lives. Likewise, in mathematics, PISA assesses how well students analyze, reason, and interpret mathematical problems in a variety of situations. Scores on the PISA scales represent skill levels along a continuum of literacy skills. PISA provides ranges of proficiency levels associated with scores that describe what a student can typically do at each level (OECD 2006).

The assessment of 15-year-old students allows countries to compare outcomes of learning as students near the end of compulsory schooling. PISA's goal is to answer the question "What knowledge and skills do students have at age 15?" In this way, PISA's achievement scores represent a "yield" of learning at age 15, rather than a direct measure of attained curriculum knowledge at a particular grade level. Fifteen-year-old students participating in PISA from the United States and other countries are drawn from a range of grade levels. Sixty-nine percent of the U.S. students were enrolled in grade 10, and another 20 percent were enrolled in grade 11 (table 2).

In addition to participating in PISA, the United States has for many years conducted assessments of student achievement at a variety of grade levels and in a variety of subject areas through the National Assessment of Educational Progress (NAEP), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS). These studies differ from PISA in terms of their purpose and design (see appendix D). NAEP reports information on the achievement of U.S. students using nationally established benchmarks of performance (i.e., *basic, proficient,* and *advanced*), based on the collaborative input of a wide range of experts and participants from government, education, business, and public sectors in the United States. Furthermore, the information is used to monitor progress in achievement over time, specific to U.S. students.

To provide a critical external perspective on the mathematics, science, and reading achievement of U.S. students, the United States participates in PISA as well as TIMSS and PIRLS. TIMSS provides the United States with information on the mathematics and science achievement of 4th- and 8th-grade U.S. students compared to students in other countries. PIRLS allows the United States to make international comparisons of the reading achievement of students in the fourth grade. TIMSS and PIRLS seek to measure students' mastery of specific knowledge, skills, and concepts and are designed to broadly reflect curricula in the United States and other participating countries; in contrast, PISA does not focus explicitly on curricular outcomes but rather on the application of knowledge to problems in a real-life context.

Table 2. Percentage distribution of U.S. 15-year-old students, by grade level: 2009

Grade level	Percent	*s.e.*
Grade 7	#	†
Grade 8	‡	†
Grade 9	10.9	0.77
Grade 10	68.5	0.98
Grade 11	20.3	0.73
Grade 12	0.1!	0.06
Total	100.0	†

† Not applicable.
# Rounds to zero.
! Interpret data with caution. Estimate is unstable due to high coefficient of variation.
‡ Reporting standards not met.
NOTE: Detail may not sum to totals because of rounding. Standard error is denoted by *s.e.*
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

How PISA 2009 Was Conducted

PISA 2009 was coordinated by the OECD and implemented at the international level by the PISA Consortium, led by the Australian Council for Educational Research (ACER).[2] The National Center for Education Statistics (NCES) of the Institute of Education Sciences (IES) at the U.S. Department of Education was responsible for the implementation of PISA in the United States. Data collection and associated tasks in the United States were carried out through a contract with Windwalker Corporation and its two subcontractors, Westat and Pearson. A steering committee (see appendix C for a list of members) provided input on the development and dissemination of PISA in the United States.

PISA 2009 was a 2-hour paper-and-pencil assessment of 15-year-olds collected from nationally representative samples of students in participating countries.[3] Like other large-scale assessments, PISA was not designed to provide individual student scores, but rather national and group estimates of performance. In PISA 2009, although each student was administered one test booklet, there were 13 test booklets in total. Each test booklet included either reading items only; reading and mathematics items; reading and science items; or reading, mathematics, and science items. As such, all students answered reading items, but not every student answered mathematics and science items (for more information on the PISA 2009 design, see the technical notes in appendix B).

PISA 2009 was administered in the United States between September and November 2009. The U.S. sample included both public and private schools, randomly selected and weighted to be representative of the nation.[4] In total, 165 schools and 5,233 students participated in PISA 2009 in the United States. The overall weighted school response rate was 68 percent before the use of replacement schools and 78 percent after the addition of replacement schools. The final weighted student response rate was 87 percent (see the technical notes in appendix B for additional details on sampling, administration, response rates, and other issues).

This report provides results for the United States in relation to the other countries participating in PISA 2009, distinguishing OECD countries and non-OECD countries and other education systems. Differences described in this report have been tested for statistical significance at the .05 level, with no adjustments for multiple comparisons. Additional information on the statistical procedures used in this report is provided in the technical notes in appendix B. For further results from PISA 2009, see the OECD publications *PISA 2009 Results (Volumes I-V)* (OECD 2010a, 2010b, 2010c, 2010d, 2010e) and the NCES website at http://nces.ed.gov/surveys/pisa.

[2] The other members of the PISA Consortium are Analyse des systèmes et des pratiques d'enseignement (aSPe, Belgium), cApStAn Linguistic Quality Control (Belgium), the German Institute for International Educational Research (DIPF), Educational Testing Service (ETS, United States), Institutt for Laererutdanning og Skoleu tvikling (ILS, Norway), Leibniz Institute for Science and Mathematics Education (IPN, Germany), the National Institute for Educational Policy Research (NIER, Japan), CRP Henri Tudor and Université de Luxembourg – EMACS (Luxembourg), and Westat (United States).

[3] Some countries also administered the PISA Electronic Reading Assessment, which was analyzed and reported separately from the paper-and-pencil assessment. The United States did not administer this optional component.

[4] The sampling data for public schools were obtained from the 2005–06 Common Core of Data (CCD), and the sampling data for private schools were obtained from the 2005–06 Private School Universe Survey (PSS).

Page intentionally left blank

U.S. Performance in Reading Literacy

PISA's major focus in 2009 was reading literacy, which is defined as follows:

> *Reading literacy is understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society* (OECD 2009, p. 23).

In assessing students' reading literacy, PISA measures the extent to which students can construct, extend, and reflect on the meaning of what they have read across a wide variety of texts associated with a wide variety of situations.

The PISA reading literacy assessment is built on three major task characteristics: "situation – the range of broad contexts or purposes for which reading takes place; text – the range of material that is read; and aspect – the cognitive approach that determines how readers engage with a text" (OECD 2009, p. 25). Text types include prose texts (such as stories, articles, and manuals) and noncontinuous texts (such as forms and advertisements) that reflect various uses or situations for which texts were constructed or the context in which knowledge and skills are applied. Reading aspects, or processes, include retrieving information; forming a broad understanding; developing an interpretation; reflecting on and evaluating the content of a text; and reflecting on and evaluating the form of a text. Sample reading literacy tasks are shown in appendix A.

Since reading literacy was the major subject area for the 2009 cycle of PISA, results are shown for the combined reading literacy scale, as well as for the three reading literacy subscales that reflect the reading aspects or processes: accessing and retrieving information, integrating and interpreting, and reflecting and evaluating. Scores on the reading literacy scale (combined and subscales) range from 0 to 1,000.[5]

[5] The reading literacy scale was established in PISA 2000 to have a mean of 500 and a standard deviation of 100. The combined reading literacy scale is made up of all items in the three subscales. However, the combined reading scale and the three subscales are each computed separately through Item Response Theory (IRT) models. Therefore, the combined reading scale score is not the average of the three subscale scores.

Performance of Students Overall

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 (table 3). Among the 33 other OECD countries, 6 countries had higher average scores than the United States, 13 had lower average scores, and 14 had average scores not measurably different from the U.S. average. Among the 64 other OECD countries, non-OECD countries, and other education systems, 9 had higher average scores than the United States, 39 had lower average scores, and 16 had average scores not measurably different from the U.S. average.

On the *reflect and evaluate* subscale, U.S. 15-year-olds had a higher average score than the OECD average (512 versus 494). The U.S. average was lower than that of 5 OECD countries and higher than that of 23 OECD countries; it was lower than that of 8 countries and other education systems and higher than that of 51 countries and other education systems overall. On the other two subscales—*access and retrieve* and *integrate and interpret*—the U.S. average was not measurably different from the OECD average (492 versus 495 and 495 versus 493, respectively).

Performance at PISA Proficiency Levels

In addition to reporting performance in terms of scale scores, PISA reports results in terms of the percentage of students at each of several proficiency levels. PISA's seven reading literacy proficiency levels, ranging from 1b to 6, are described in exhibit 1 (see appendix B for information about how the proficiency levels are created).

Table 3. Average scores of 15-year-old students on combined reading literacy scale and reading literacy subscales, by country: 2009

		Reading literacy subscales					
Combined reading literacy scale		Access and retrieve		Integrate and interpret		Reflect and evaluate	
Country	Score	Country	Score	Country	Score	Country	Score
OECD average	493	OECD average	495	OECD average	493	OECD average	494
OECD countries		*OECD countries*		*OECD countries*		*OECD countries*	
Korea, Republic of	539	Korea, Republic of	542	Korea, Republic of	541	Korea, Republic of	542
Finland	536	Finland	532	Finland	538	Finland	536
Canada	524	Japan	530	Canada	522	Canada	535
New Zealand	521	New Zealand	521	Japan	520	New Zealand	531
Japan	520	Netherlands	519	New Zealand	517	Australia	523
Australia	515	Canada	517	Australia	513	Japan	521
Netherlands	508	Belgium	513	Netherlands	504	**United States**	**512**
Belgium	506	Australia	513	Belgium	504	Netherlands	510
Norway	503	Norway	512	Poland	503	Belgium	505
Estonia	501	Iceland	507	Iceland	503	Norway	505
Switzerland	501	Switzerland	505	Norway	502	United Kingdom	503
Poland	500	Sweden	505	Switzerland	502	Estonia	503
Iceland	500	Estonia	503	Germany	501	Ireland	502
United States	**500**	Denmark	502	Estonia	500	Sweden	502
Sweden	497	Hungary	501	France	497	Poland	498
Germany	497	Germany	501	Hungary	496	Switzerland	497
Ireland	496	Poland	500	**United States**	**495**	Portugal	496
France	496	Ireland	498	Sweden	494	Iceland	496
Denmark	495	**United States**	**492**	Ireland	494	France	495
United Kingdom	494	France	492	Denmark	492	Denmark	493
Hungary	494	United Kingdom	491	United Kingdom	491	Germany	491
Portugal	489	Slovak Republic	491	Italy	490	Greece	489
Italy	486	Slovenia	489	Slovenia	489	Hungary	489
Slovenia	483	Portugal	488	Czech Republic	488	Spain	483
Greece	483	Italy	482	Portugal	487	Israel	483
Spain	481	Spain	480	Greece	484	Italy	482
Czech Republic	478	Czech Republic	479	Slovak Republic	481	Turkey	473
Slovak Republic	477	Austria	477	Spain	481	Luxembourg	471
Israel	474	Luxembourg	471	Luxembourg	475	Slovenia	470
Luxembourg	472	Greece	468	Israel	473	Slovak Republic	466
Austria	470	Turkey	467	Austria	471	Austria	463
Turkey	464	Israel	463	Turkey	459	Czech Republic	462
Chile	449	Chile	444	Chile	452	Chile	452
Mexico	425	Mexico	433	Mexico	418	Mexico	432

- Average is higher than the U.S. average
- Average is not measurably different from the U.S. average
- Average is lower than the U.S. average

See notes at end of table.

Table 3. Average scores of 15-year-old students on combined reading literacy scale and reading literacy subscales, by country: 2009–Continued

Combined reading literacy scale		Reading literacy subscales					
		Access and retrieve		Integrate and interpret		Reflect and evaluate	
Country	Score	Country	Score	Country	Score	Country	Score
Non-OECD countries		***Non-OECD countries***		***Non-OECD countries***		***Non-OECD countries***	
Shanghai-China	556	*Shanghai-China*	549	*Shanghai-China*	558	*Shanghai-China*	557
Hong Kong-China	533	*Hong Kong-China*	530	*Hong Kong-China*	530	*Hong Kong-China*	540
Singapore	526	Singapore	526	Singapore	525	Singapore	529
Liechtenstein	499	Liechtenstein	508	*Chinese Taipei*	499	Liechtenstein	498
Chinese Taipei	495	*Chinese Taipei*	496	Liechtenstein	498	*Chinese Taipei*	493
Macao-China	487	*Macao-China*	493	*Macao-China*	488	Latvia	492
Latvia	484	Croatia	492	Latvia	484	*Macao-China*	481
Croatia	476	Lithuania	476	Croatia	472	Croatia	471
Lithuania	468	Latvia	476	Lithuania	469	*Dubai-UAE*	466
Dubai-UAE	459	Russian Federation	469	Russian Federation	467	Lithuania	463
Russian Federation	459	*Dubai-UAE*	458	*Dubai-UAE*	457	Russian Federation	441
Serbia, Republic of	442	Serbia, Republic of	449	Serbia, Republic of	445	Uruguay	436
Bulgaria	429	Thailand	431	Bulgaria	436	Serbia, Republic of	430
Uruguay	426	Bulgaria	430	Romania	425	Tunisia	427
Romania	424	Uruguay	424	Uruguay	423	Romania	426
Thailand	421	Romania	423	Montenegro, Republic of	420	Brazil	424
Trinidad and Tobago	416	Trinidad and Tobago	413	Trinidad and Tobago	419	Colombia	422
Colombia	413	Montenegro, Republic of	408	Thailand	416	Thailand	420
Brazil	412	Brazil	407	Colombia	411	Bulgaria	417
Montenegro, Republic of	408	Colombia	404	Jordan	410	Trinidad and Tobago	413
Jordan	405	Indonesia	399	Brazil	406	Indonesia	409
Tunisia	404	Kazakhstan	397	Argentina	398	Jordan	407
Indonesia	402	Argentina	394	Indonesia	397	Argentina	402
Argentina	398	Jordan	394	Kazakhstan	397	Montenegro, Republic of	383
Kazakhstan	390	Tunisia	393	Tunisia	393	Panama	377
Albania	385	Albania	380	Albania	393	Albania	376
Qatar	372	Peru	364	Qatar	379	Qatar	376
Panama	371	Panama	363	Azerbaijan	373	Kazakhstan	373
Peru	370	Azerbaijan	361	Panama	372	Peru	368
Azerbaijan	362	Qatar	354	Peru	371	Azerbaijan	335
Kyrgyz Republic	314	Kyrgyz Republic	299	Kyrgyz Republic	327	Kyrgyz Republic	300

- Average is higher than the U.S. average
- Average is not measurably different from the U.S. average
- Average is lower than the U.S. average

NOTE: The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Because the Program for International Student Assessment (PISA) is principally an OECD study, the results for non-OECD countries are displayed separately from those of the OECD countries and are not included in the OECD average. Countries are ordered on the basis of average scores, from highest to lowest within the OECD countries and non-OECD countries. Scores are reported on a scale from 0 to 1,000. Score differences as noted between the United States and other countries (as well as between the United States and the OECD average) are significantly different at the .05 level of statistical significance. The standard errors of the estimates are shown in table R1 available at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp. 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), 2009.

Exhibit 1. Description of PISA proficiency levels on combined reading literacy scale: 2009

Proficiency level and lower cut point score	Task descriptions
Level 6 698	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. Scores are reported on a scale from 0 to 1,000.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

In reading literacy, 30 percent[6] of U.S. students scored at or above proficiency level 4, that is, at levels 4, 5, or 6, as shown in figure 3. Level 4 is the level at which students are "capable of difficult reading tasks, such as locating embedded information, construing meaning from nuances of language and critically evaluating a text" (OECD 2010a, p. 51). At levels 5 and 6 students demonstrate higher-level reading skills and may be referred to as "top performers" in reading. While there was no measurable difference between the percentage of U.S. students and the percentage of students in the OECD countries on average who performed at or above level 4, a higher percentage of U.S. students performed at level 5 than the OECD average (8 versus 7 percent). In comparison to the United States, 7 OECD countries and 3 non-OECD countries and other education systems had higher percentages of students who performed at or above level 4 in reading literacy; 14 OECD countries and 27 non-OECD countries and other education systems had lower percentages of students who performed at or above level 4; and for 12 OECD countries and 1 non-OECD country, there were no measurable differences in the percentages of students who performed at or above level 4 (data shown in table R7A at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp).

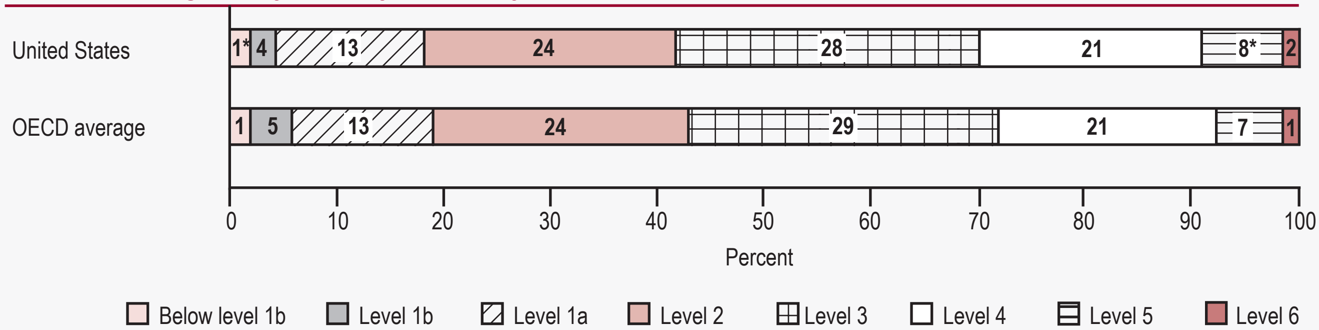
Eighteen percent of U.S. students scored below level 2 (that is, at levels 1a or 1b or below 1b). Students performing below level 2 are below what OECD calls "a baseline level of proficiency, at which students begin to demonstrate the reading literacy competencies that will enable them to participate effectively and productively in life" (OECD 2010a, p. 52). Students performing at levels 1a and 1b are able to perform only the least complex reading tasks on the PISA assessment such as locating explicitly stated information in the text and making simple connections between text and common knowledge (level 1a) or doing so in simple texts (level 1b), as described in exhibit 1. Students below level 1b are not able to routinely perform these tasks; this does not mean that they have no literacy skills but the PISA assessment cannot accurately characterize their skills. There was no measurable difference between the percentage of U.S. students and the percentage of students in the OECD countries on average demonstrating proficiency below level 2.

Differences in Performance by Selected Student and School Characteristics

This section reports performance on the PISA combined reading literacy scale by selected characteristics of students: sex, racial/ethnic background, and the socioeconomic context of their schools. The results cannot be used to demonstrate a cause-and-effect relationship between these variables and student performance. Student performance can be affected by a complex mix of educational and other factors that are not accounted for in these analyses.

[6] This estimate was calculated using unrounded percentages at levels 4, 5, and 6.

Figure 3. Percentage distribution of 15-year-old students in the United States and OECD countries on combined reading literacy scale, by proficiency level: 2009



**p* < .05. Significantly different from the corresponding OECD average percentage at the .05 level of statistical significance.
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. Exact cut point scores are as follows: below level 1b (a score less than or equal to 262.04); level 1b (a score greater than 262.04 and less than or equal to 334.75); level 1a (a score greater than 334.75 and less than or equal to 407.47); level 2 (a score greater than 407.47 and less than or equal to 480.18); level 3 (a score greater than 480.18 and less than or equal to 552.89); level 4 (a score greater than 552.89 and less than or equal to 625.61); level 5 (a score greater than 625.61 and less than or equal to 698.32); and level 6 (a score greater than 698.32). The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. Detail may not sum to totals because of rounding. The standard errors of the estimates are shown in table R7 available at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

Sex

Female students scored higher, on average, than male students on the combined reading literacy scale in all 65 participating countries and other education systems (table 4). The gender gap ranged from a difference of 9 scale score points in Colombia to 62 scale score points in Albania.

In the United States, the difference (25 scale score points) was smaller than the difference in the OECD countries, on average (39 scale score points), and smaller than the differences in 45 countries and other education systems (24 OECD countries and 21 non-OECD countries and other education systems).

Table 4. Average scores of 15-year-old female and male students on combined reading literacy scale, by country: 2009

	Female		Male		Female-male difference	
Country	Score	s.e.	Score	s.e.	Score difference*	s.e.
OECD average	513	0.5	474	0.6	39	0.6
OECD countries						
Chile	461	3.6	439	3.9	22	4.1
Netherlands	521	5.3	496	5.1	24	2.4
United States	**513**	**3.8**	**488**	**4.2**	**25**	**3.4**
Mexico	438	2.1	413	2.1	25	1.6
United Kingdom	507	2.9	481	3.5	25	4.5
Belgium	520	2.9	493	3.4	27	4.4
Denmark	509	2.5	480	2.5	29	2.9
Spain	496	2.2	467	2.2	29	2.0
Canada	542	1.7	507	1.8	34	1.9
Korea, Republic of	558	3.8	523	4.9	35	5.9
Australia	533	2.6	496	2.9	37	3.1
Hungary	513	3.6	475	3.9	38	4.0
Portugal	508	2.9	470	3.5	38	2.4
Switzerland	520	2.7	481	2.9	39	2.5
Japan	540	3.7	501	5.6	39	6.8
Ireland	515	3.1	476	4.2	39	4.7
Luxembourg	492	1.5	453	1.9	39	2.3
Germany	518	2.9	478	3.6	40	3.9
France	515	3.4	475	4.3	40	3.7
Austria	490	4.0	449	3.8	41	5.5
Israel	495	3.4	452	5.2	42	5.2
Turkey	486	4.1	443	3.7	43	3.7
Iceland	522	1.9	478	2.1	44	2.8
Estonia	524	2.8	480	2.9	44	2.5
Sweden	521	3.1	475	3.2	46	2.7
New Zealand	544	2.6	499	3.6	46	4.3
Italy	510	1.9	464	2.3	46	2.8
Greece	506	3.5	459	5.5	47	4.3
Norway	527	2.9	480	3.0	47	2.9
Czech Republic	504	3.0	456	3.7	48	4.1
Poland	525	2.9	476	2.8	50	2.5
Slovak Republic	503	2.8	452	3.5	51	3.5
Slovenia	511	1.4	456	1.6	55	2.3
Finland	563	2.4	508	2.6	55	2.3

■ Female-male difference is smaller than the U.S. difference

□ Female-male difference is not measurably different from the U.S. difference

■ Female-male difference is larger than the U.S. difference

See notes at end of table.

Table 4. Average scores of 15-year-old female and male students on combined reading literacy scale, by country: 2009—Continued

Country	Female		Male		Female-male difference	
	Score	*s.e.*	Score	*s.e.*	Score difference*	*s.e.*
Non-OECD countries						
Colombia	418	4.0	408	4.5	9!	3.8
Peru	381	4.9	359	4.2	22	4.7
Azerbaijan	374	3.3	350	3.7	24	2.4
Brazil	425	2.8	397	2.9	29	1.7
Tunisia	418	3.0	387	3.2	31	2.2
Singapore	542	1.5	511	1.7	31	2.3
Liechtenstein	516	4.5	484	4.5	32	7.1
Hong Kong-China	550	2.8	518	3.3	33	4.4
Panama	387	7.3	354	7.0	33	6.7
Macao-China	504	1.2	470	1.3	34	1.7
Indonesia	420	3.9	383	3.8	37	3.3
Argentina	415	4.9	379	5.1	37	3.8
Chinese Taipei	514	3.6	477	3.7	37	5.3
Thailand	438	3.1	400	3.3	38	3.8
Serbia, Republic of	462	2.5	422	3.3	39	3.0
Shanghai-China	576	2.3	536	3.0	40	2.9
Uruguay	445	2.8	404	3.2	42	3.1
Romania	445	4.3	403	4.6	43	4.4
Kazakhstan	412	3.4	369	3.2	43	2.7
Russian Federation	482	3.4	437	3.6	45	2.7
Latvia	507	3.1	460	3.4	47	3.2
Qatar	397	1.0	347	1.3	50	1.8
Dubai-UAE	485	1.5	435	1.7	51	2.3
Croatia	503	3.7	452	3.4	51	4.6
Montenegro, Republic of	434	2.1	382	2.1	53	2.6
Kyrgyz Republic	340	3.2	287	3.8	53	2.7
Jordan	434	4.1	377	4.7	57	6.2
Trinidad and Tobago	445	1.6	387	1.9	58	2.5
Lithuania	498	2.6	439	2.8	59	2.8
Bulgaria	461	5.8	400	7.3	61	4.7
Albania	417	3.9	355	5.1	62	4.4

■ Female-male difference is smaller than the U.S. difference

□ Female-male difference is not measurably different from the U.S. difference

■ Female-male difference is larger than the U.S. difference

! Interpret data with caution. Estimate is unstable due to high coefficient of variation.

* $p < .05$. All differences between females and males are significantly different at the .05 level of statistical significance. Differences were computed using unrounded numbers.

NOTE: The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Because the Program for International Student Assessment (PISA) is principally an OECD study, the results for non-OECD countries are displayed separately from those of the OECD countries. Scores are reported on a scale from 0 to 1,000. Standard error is noted by *s.e.* 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), 2009.

Race/Ethnicity

Racial and ethnic groups vary by country, so it is not possible to compare performance of students in individual countries by students' race/ethnicity. Therefore, only results for the United States are presented.

On the combined reading literacy scale, White (non-Hispanic) and Asian (non-Hispanic) students had higher average scores (525 and 541, respectively) than the overall OECD and U.S. average scores, while Black (non-Hispanic) and Hispanic students had lower average scores (441 and 466, respectively) than the overall OECD and U.S. average scores (table 5). The average scores of students who reported two or more races (502) were not measurably different from the overall OECD or U.S. average scores.

The average scores of White (non-Hispanic) students, Asian (non-Hispanic) students, and students who reported two or more races (525, 541, and 502, respectively) were in the range of PISA's proficiency level 3 (signifies a score of greater than 480 and less than or equal to 553), while the average scores of Black (non-Hispanic) and Hispanic students (441 and 466, respectively) were in the range of PISA's proficiency level 2 (signifies a score of greater than 407 and less than or equal to 480). These findings describe average performance and do not describe variation within the subgroup. Students at level 3 on the reading literacy scale are typically successful at "reading tasks of moderate complexity, such as locating multiple pieces of information, making links between different parts of a text, and relating it to familiar everyday knowledge," as described in exhibit 1, and other tasks that might be expected to be commonly demanded of young and older adults across OECD countries in their everyday lives (OECD 2010a, p. 51). At level 2, which "can be considered a baseline level of proficiency, at which students begin to demonstrate the reading literacy competencies that will enable them to participate effectively and productively in life" (OECD 2010a, p. 52), students can typically locate information that meets several conditions, make comparisons or contrasts around a single feature, determine what a well-defined part of a text means even when the information is not prominent, and make connections between the text and personal experience.

Table 5. Average scores of U.S. 15-year-old students on combined reading literacy scale, by race/ethnicity: 2009

Race/ethnicity	Score	*s.e.*
U.S. average	**500**	**3.7**
White, non-Hispanic	525*	3.8
Black, non-Hispanic	441*	7.2
Hispanic	466*	4.3
Asian, non-Hispanic	541*	9.4
American Indian/Alaska Native, non-Hispanic	‡	†
Native Hawaiian/Other Pacific Islander, non-Hispanic	‡	†
Two or more races, non-Hispanic	502	6.4
OECD average	**493**	**0.5**

† Not applicable.
‡ Reporting standards not met.
*$p < .05$. Significantly different from the U.S. and OECD averages at the .05 level of statistical significance.
NOTE: Black includes African American, and Hispanic includes Latino. Students who identified themselves as being of Hispanic origin were classified as Hispanic, regardless of their race. Although data for some race/ethnicities are not shown separately because the reporting standards were not met, they are included in the U.S. totals shown throughout the report. The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. Standard error is noted by *s.e.*
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

School Socioeconomic Contexts

The percentage of students in a school who are eligible for free or reduced-price lunch (FRPL-eligible) through the National School Lunch Program is an indicator, in the United States, of the socioeconomic status of families served by the school. Other countries have different indicators of school socioeconomic context and thus only results for the United States are shown by the percentage of students in schools who are FRPL-eligible. Data are for public schools only.

Students in public schools in which half or more of students were eligible for free or reduced-price lunch (50 to 74.9 percent and 75 percent or more) scored, on average, below the overall OECD and U.S. average scores (table 6). Students in schools in which less than 25 percent of students were FRPL-eligible (10 to 24.9 percent and less than 10 percent) scored, on average, above the overall OECD and U.S. average scores. The average scores of students in schools in which 25 to 49.9 percent were FRPL-eligible were above the overall OECD average but not measurably different from the U.S. average.

The average scale score of students in schools with less than 10 percent of FRPL-eligible students (551) was at the upper end of proficiency level 3 (upper cut point is 553), while students in schools with 75 percent or more of FRPL-eligible students performed at the middle of level 2, with an average scale score of 446 (level 2 midpoint is 444), a difference of 105 scale score points.

Table 6. 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: 2009

Percent of students eligible for free or reduced-price lunch	Score	*s.e.*
U.S. average	**500**	**3.7**
Less than 10 percent	551*	7.6
10 to 24.9 percent	527*	6.5
25 to 49.9 percent	502**	4.1
50 to 74.9 percent	471*	6.5
75 percent or more	446*	6.9
OECD average	**493**	**0.5**

*$p < .05$. Significantly different from the U.S. and OECD averages at the .05 level of statistical significance.
** $p < .05$. Significantly different from the OECD average at the .05 level of statistical significance, but not significantly different from the U.S. average.
NOTE: The National School Lunch Program provides free or reduced-price lunch for students meeting certain income guidelines. The percentage of students receiving such lunch is an indicator of the socioeconomic level of families served by the school. The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. Standard error is noted by *s.e.* Data are for public schools only.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009

Trends in Average Performance

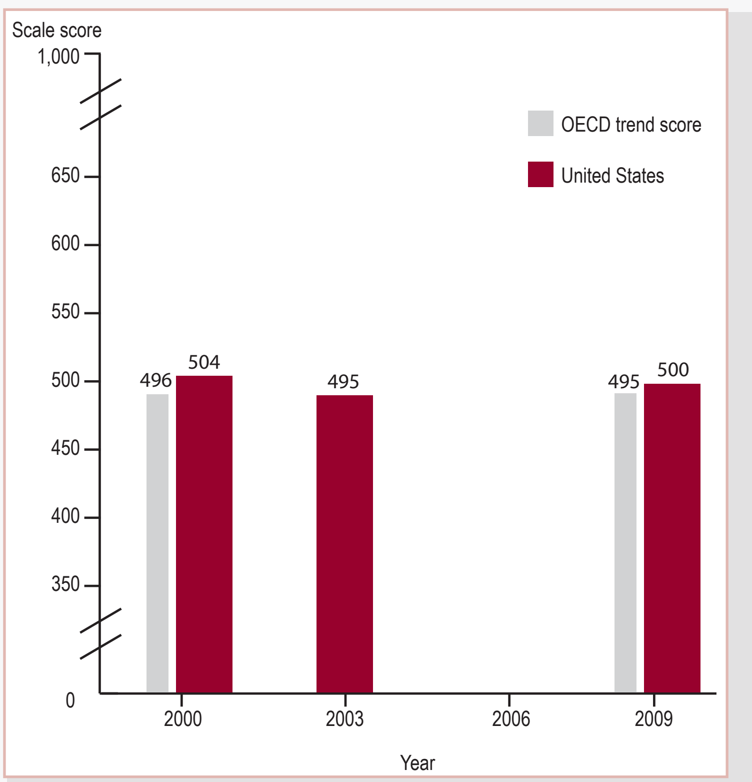
There was no measurable difference between the average score of U.S. students in reading literacy in 2000 (504), the last time in which reading literacy was the major domain assessed in PISA, and 2009 (500), or between 2003 (495) and 2009 (figure 4).[7] There also were no measurable differences between the U.S. average score and the OECD average score in 2000 or in 2009 when the OECD averages were 496 and 495, respectively.

The PISA 2000 and 2009 OECD averages used in the analysis of trends in reading literacy over time are based on the averages of the 27 OECD countries with comparable data for 2000 and 2009.[8] As a result, the reading literacy OECD average score for PISA 2000 differs from previously published reports and the reading literacy OECD average score for PISA 2009 differs from that reported in other tables in this report. The recalculated OECD averages are referred to as OECD trend scores. The U.S. averages in 2000 and 2009 are compared with OECD trend scores in 2000 and 2009 because reading literacy was the major domain assessed in those years.

[7] U.S. reading results for PISA 2006 are not available due to a printing error in the U.S. test booklets in 2006.

[8] OECD trend scores are not reported for 2003 and 2006 because data were not available for all 27 comparable countries. The seven current OECD members not included in the OECD averages used to report on trends in reading literacy include the Slovak Republic and Turkey, which joined PISA in 2003; Estonia and Slovenia, which joined PISA in 2006; Luxembourg, which experienced substantial changes in its assessment conditions between 2000 and 2003; and the Netherlands and the United Kingdom, which did not meet the PISA response-rate standards in 2000. The OECD excluded the data for Austria from the trend analysis in its report (OECD 2010e) because of a concern over a data collection issue in 2009; however, after consultation with Austrian officials, NCES kept the Austrian data in the U.S. trend reporting.

Figure 4. Average scores of 15-year-old students in the United States and OECD countries on reading literacy scale: 2000, 2003, and 2009



NOTE: PISA 2006 reading literacy results are not reported for the United States because of an error in printing the test booklets. For more details, see Baldi et al. 2007 (available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008016). The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. There were no statistically significant differences between the U.S. average score and the OECD average score in 2000 or in 2009. The standard errors of the estimates are shown in table R5 available at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2000, 2003, and 2009.

U.S. Performance in Mathematics Literacy

In PISA 2009, mathematics literacy is defined as follows:

> *An individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen* (OECD 2009, p. 84).

Performance of Students Overall

U.S. 15-year-olds had an average score of 487 on the mathematics literacy scale, which was lower than the OECD average score of 496 (table 7).[9] 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. Among the 64 other OECD countries, non-OECD countries, and other education systems, 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.

[9] The mathematics literacy scale was established in PISA 2003 to have a mean of 500 and a standard deviation of 100.

U.S. Performance in Mathematics Literacy

Table 7. Average scores of 15-year-old students on mathematics literacy scale, by country: 2009

Mathematics literacy scale	
Country	**Score**
OECD average	496
OECD countries	
Korea, Republic of	546
Finland	541
Switzerland	534
Japan	529
Canada	527
Netherlands	526
New Zealand	519
Belgium	515
Australia	514
Germany	513
Estonia	512
Iceland	507
Denmark	503
Slovenia	501
Norway	498
France	497
Slovak Republic	497
Austria	496
Poland	495
Sweden	494
Czech Republic	493
United Kingdom	492
Hungary	490
Luxembourg	489
United States	**487**
Ireland	487
Portugal	487
Spain	483
Italy	483
Greece	466
Israel	447
Turkey	445
Chile	421
Mexico	419

Mathematics literacy scale	
Country	**Score**
Non-OECD countries	
Shanghai-China	600
Singapore	562
Hong Kong-China	555
Chinese Taipei	543
Liechtenstein	536
Macao-China	525
Latvia	482
Lithuania	477
Russian Federation	468
Croatia	460
Dubai-UAE	453
Serbia, Republic of	442
Azerbaijan	431
Bulgaria	428
Romania	427
Uruguay	427
Thailand	419
Trinidad and Tobago	414
Kazakhstan	405
Montenegro, Republic of	403
Argentina	388
Jordan	387
Brazil	386
Colombia	381
Albania	377
Tunisia	371
Indonesia	371
Qatar	368
Peru	365
Panama	360
Kyrgyz Republic	331

- Average is higher than the U.S. average
- Average is not measurably different from the U.S. average
- Average is lower than the U.S. average

NOTE: The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Because the Program for International Student Assessment (PISA) is principally an OECD study, the results for non-OECD countries are displayed separately from those of the OECD countries and are not included in the OECD average. Countries are ordered on the basis of average scores, from highest to lowest within the OECD countries and non-OECD countries. Scores are reported on a scale from 0 to 1,000. Score differences as noted between the United States and other countries (as well as between the United States and the OECD average) are significantly different at the .05 level of statistical significance. The standard errors of the estimates are shown in table M1 available at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp. 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), 2009.

Performance at PISA Proficiency Levels

PISA's six mathematics literacy proficiency levels, ranging from 1 to 6, are described in exhibit 2 (see appendix B for information about how the proficiency levels are created).

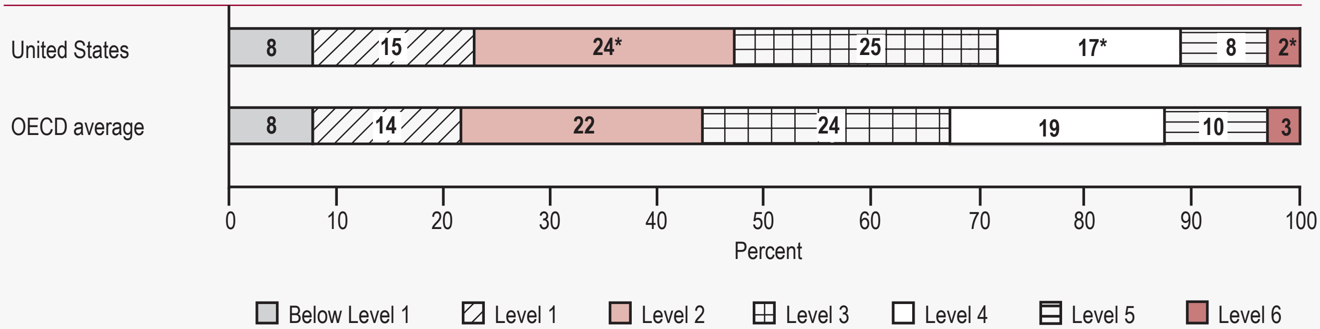
Exhibit 2. Description of PISA proficiency levels on mathematics literacy scale: 2009

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. Scores are reported on a scale from 0 to 1,000.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

In mathematics literacy, 27 percent of U.S. students scored at or above proficiency level 4, that is, at levels 4, 5, or 6 (figure 5 and exhibit 2). This is lower than the 32 percent of students in the OECD countries on average that scored at or above level 4. Level 4 is the level at which students can complete higher order tasks such as "solv[ing] problems that involve visual and spatial reasoning...in unfamiliar contexts" and "carry[ing] out sequential processes" (OECD 2004, p. 55). A lower percentage of U.S. students performed at level 4 than the OECD average (17 percent versus 19 percent) and at level 6 (2 percent versus 3 percent). Twenty-three percent of U.S. students scored below level 2 (that is, at level 1 or below level 1), what OECD calls a "a baseline level of mathematics proficiency on the PISA scale at which students begin to demonstrate the kind of literacy skills that enable them to actively use mathematics" (OECD 2004, p. 56). There was no measurable difference between the percentage of U.S. students and the percentage of students in the OECD countries on average demonstrating proficiency below level 2. A description of the general competencies and tasks 15-year-old students typically can do, by proficiency level, for the mathematics literacy scale is shown in exhibit 2. In comparison to the United States, 16 OECD countries and 6 non-OECD countries and other education systems had higher percentages of students who performed at or above level 4 in mathematics literacy; 5 OECD countries and 25 non-OECD countries and other education systems had lower percentages of students who performed at or above level 4; and for 12 OECD countries, there were no measurable differences in the percentage of students who performed at or above level 4 (data shown in table M4A at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp).

Figure 5. Percentage distribution of 15-year-old students in the United States and OECD countries on mathematics literacy scale, by proficiency level: 2009



*$p < .05$. Significantly different from the corresponding OECD average percentage at the .05 level of statistical significance.

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. Exact cut point scores are as follows: below level 1 (a score less than or equal to 357.77); level 1 (a score greater than 357.77 and less than or equal to 420.07); level 2 (a score greater than 420.07 and less than or equal to 482.38); level 3 (a score greater than 482.38 and less than or equal to 544.68); level 4 (a score greater than 544.68 and less than or equal to 606.99); level 5 (a score greater than 606.99 and less than or equal to 669.30); and level 6 (a score greater than 669.30). Scores are reported on a scale from 0 to 1,000. The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Detail may not sum to totals because of rounding. The standard errors of the estimates are shown in table M4 available at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp.

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

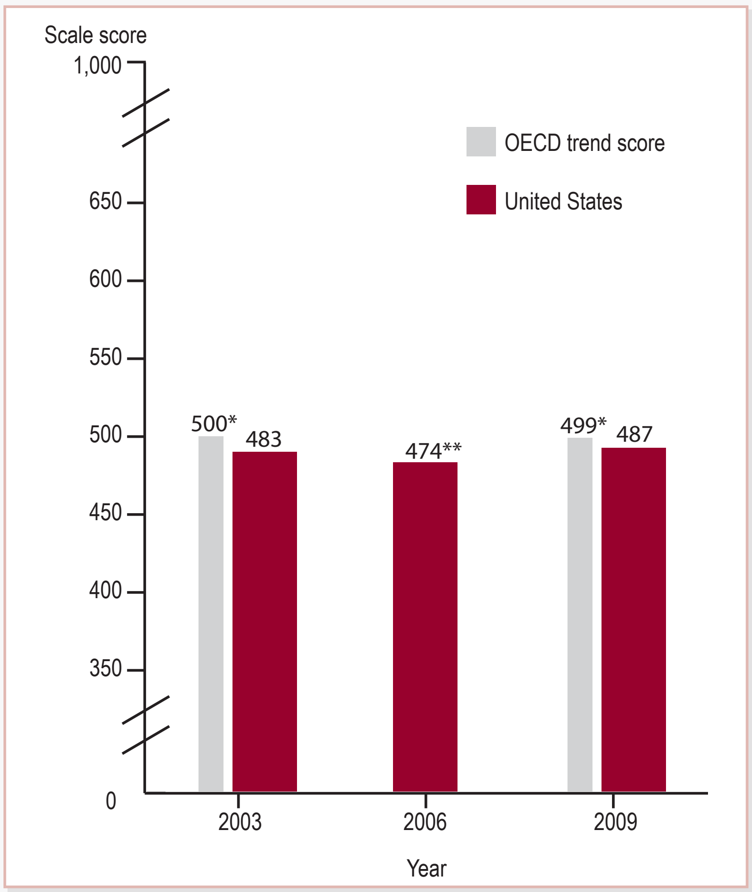
Trends in Average Performance

The U.S. average score in mathematics literacy in 2009 (487) was higher than the U.S. average in 2006 (474) but not measurably different from the U.S. average in 2003 (483), the earliest time point to which PISA 2009 performance can be compared in mathematics literacy (figure 6). U.S. students' average scores were lower than the OECD average scores in each of these years (2003 and 2009).

The PISA 2003 and 2009 OECD averages used in the analysis of trends in mathematics literacy over time are based on the averages of the 29 OECD countries with comparable data for 2003 and 2009.[10] As a result, the mathematics literacy OECD average score for PISA 2003 differs from previously published reports and the mathematics literacy OECD average score for PISA 2009 differs from that reported in other tables in this report. The recalculated OECD averages are referred to as OECD trend scores. The U.S. averages in 2003 and 2009 are compared with the OECD trend scores in 2003 and 2009 because in 2003 mathematics literacy was the major domain assessed.

[10] The OECD trend score is not reported for 2006 because data were not available for all 29 comparable countries. The five current members not included in the OECD averages used to report on trends in mathematics literacy include Chile, Estonia, Israel, and Slovenia, which did not participate in 2003, and the United Kingdom, which did not meet PISA response-rate standards for the 2003 assessment. The OECD excluded the data for Austria from the trend analysis in its report (OECD 2010e) because of a concern over a data collection issue in 2009; however, after consultation with Austrian officials, NCES kept the Austrian data in the U.S. trend reporting.

Figure 6. Average scores of 15-year-old students in the United States and OECD countries on mathematics literacy scale: 2003, 2006, and 2009



*$p < .05$. U.S. average is significantly different from the OECD trend score at the .05 level of statistical significance.
**$p < .05$. U.S. average in 2006 is significantly different from the U.S. average in 2009 at the .05 level of statistical significance.
NOTE: The PISA mathematics framework was revised in 2003. Because of changes in the framework, it is not possible to compare mathematics learning outcomes from PISA 2000 with those from PISA 2003, 2006, and 2009. For more details, see OECD (2010e). The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. The standard errors of the estimates are shown in table M2 available at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2003, 2006, and 2009.

Page intentionally left blank

U.S. Performance in Science Literacy

In PISA 2009, science literacy is defined as follows:

An individual's scientific knowledge and use of that knowledge to identify questions, to acquire new knowledge, to explain scientific phenomena, and to draw evidence based conclusions about science-related issues; understanding of the characteristic features of science as a form of human knowledge and inquiry; awareness of how science and technology shape our material, intellectual, and cultural environments; and willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen (OECD 2009, p. 128).

Performance of Students Overall

On the science literacy scale, the average score of U.S. students (502) was not measurably different from the OECD average (501) (table 8).[11] 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. Among the 64 other OECD countries, non-OECD countries, and other education systems, 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.

[11] The science literacy scale was established in PISA 2006 to have a mean of 500 and a standard deviation of 100.

Table 8. Average scores of 15-year-old students on science literacy scale, by country: 2009

Science literacy scale	
Country	Score
OECD average	501
OECD countries	
Finland	554
Japan	539
Korea, Republic of	538
New Zealand	532
Canada	529
Estonia	528
Australia	527
Netherlands	522
Germany	520
Switzerland	517
United Kingdom	514
Slovenia	512
Poland	508
Ireland	508
Belgium	507
Hungary	503
United States	**502**
Czech Republic	500
Norway	500
Denmark	499
France	498
Iceland	496
Sweden	495
Austria	494
Portugal	493
Slovak Republic	490
Italy	489
Spain	488
Luxembourg	484
Greece	470
Israel	455
Turkey	454
Chile	447
Mexico	416

Science literacy scale	
Country	Score
Non-OECD countries	
Shanghai-China	575
Hong Kong-China	549
Singapore	542
Chinese Taipei	520
Liechtenstein	520
Macao-China	511
Latvia	494
Lithuania	491
Croatia	486
Russian Federation	478
Dubai-UAE	466
Serbia, Republic of	443
Bulgaria	439
Romania	428
Uruguay	427
Thailand	425
Jordan	415
Trinidad and Tobago	410
Brazil	405
Colombia	402
Montenegro, Republic of	401
Argentina	401
Tunisia	401
Kazakhstan	400
Albania	391
Indonesia	383
Qatar	379
Panama	376
Azerbaijan	373
Peru	369
Kyrgyz Republic	330

Average is higher than the U.S. average (Finland through Slovenia; Shanghai-China through Macao-China)

Average is not measurably different from the U.S. average (OECD average; Poland through Portugal; Latvia)

Average is lower than the U.S. average (Slovak Republic through Mexico; Lithuania through Kyrgyz Republic)

NOTE: The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Because the Program for International Student Assessment (PISA) is principally an OECD study, the results for non-OECD countries are displayed separately from those of the OECD countries and are not included in the OECD average. Countries are ordered on the basis of average scores, from highest to lowest within the OECD countries and non-OECD countries. Scores are reported on a scale from 0 to 1,000. Score differences as noted between the United States and other countries (as well as between the United States and the OECD average) are significantly different at the .05 level of statistical significance. The standard errors of the estimates are shown in table S1 available at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp. 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), 2009.

Performance at PISA Proficiency Levels

PISA's six science literacy proficiency levels, ranging from 1 to 6, are described in exhibit 3 (see appendix B for information about how the proficiency levels are created).

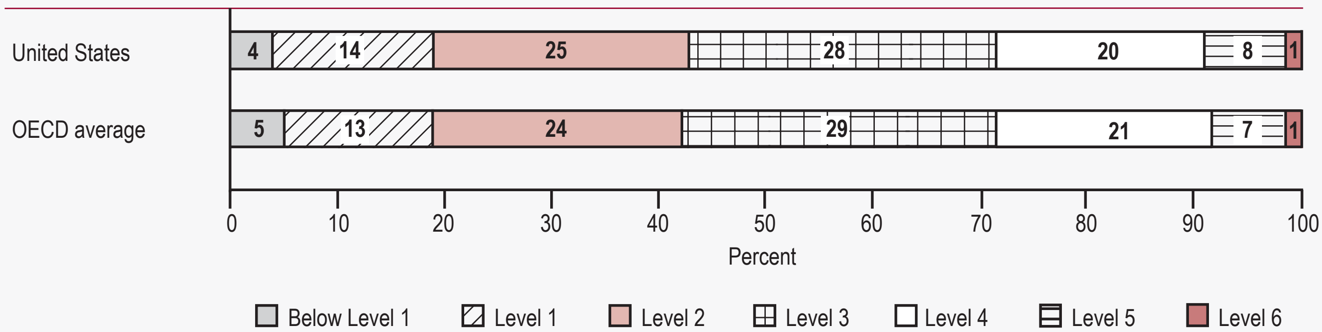
Exhibit 3. Description of PISA proficiency levels on science literacy scale: 2009

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 science literacy levels according to their scores. Cut point scores in the exhibit are rounded; exact cut point scores are provided in appendix B. Scores are reported on a scale from 0 to 1,000.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

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, that is, at levels 4, 5, or 6. Level 4 is the level at which students can complete higher order tasks such as "select[ing] and integrat[ing] explanations from different disciplines of science or technology" and "link[ing] those explanations directly to...life situations" (OECD 2007, p. 43). Eighteen percent of U.S. students and students in the OECD countries on average scored below level 2, that is, at level 1 or below level 1 (figure 7). Students performing below level 2 are below what OECD calls a "baseline level of proficiency…at which students begin to demonstrate the science competencies that will enable them to participate effectively and productively in life situations related to science and technology" (OECD 2007, p. 44). There also were no measurable differences between the percentages of U.S. students and students in the OECD countries on average that scored at the individual proficiency levels. In comparison to the United States, 13 OECD countries and 5 non-OECD countries and other education systems had higher percentages of students who performed at or above level 4 in science literacy; 11 OECD countries and 25 non-OECD countries and other education systems had lower percentages of students who performed at or above level 4; and for 9 OECD countries and 1 non-OECD education system, there were no measurable differences in the percentage of students who performed at or above level 4 (data shown in table S4A at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp).

Figure 7. Percentage distribution of 15-year-old students in the United States and OECD countries on science literacy scale, by proficiency level: 2009



NOTE: To reach a particular proficiency level, a student must correctly answer a majority of items at that level. Students were classified into science literacy levels according to their scores. Exact cut point scores are as follows: below level 1 (a score less than or equal to 334.94); level 1 (a score greater than 334.94 and less than or equal to 409.54); level 2 (a score greater than 409.54 and less than or equal to 484.14); level 3 (a score greater than 484.14 and less than or equal to 558.73); level 4 (a score greater than 558.73 and less than or equal to 633.33); level 5 (a score greater than 633.33 and less than or equal to 707.93); and level 6 (a score greater than 707.93). Scores are reported on a scale from 0 to 1,000. The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Detail may not sum to totals because of rounding. There were no statistically significant differences between U.S. students and the OECD average in the percentages of students at each proficiency level. The standard errors of the estimates are shown in table S4 available at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

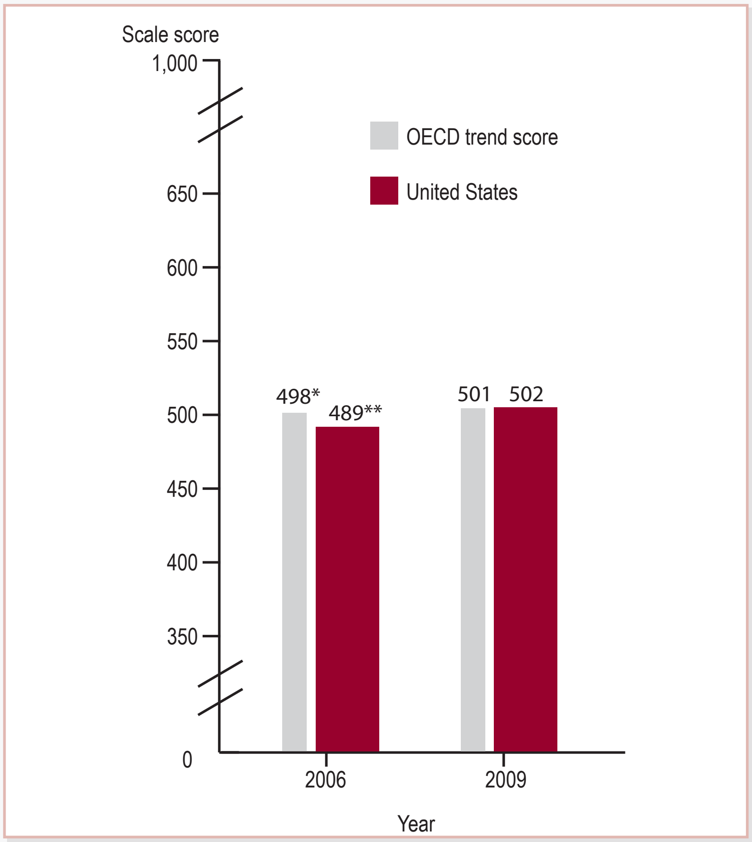
Trends in Average Performance

The U.S. average score in science literacy in 2009 (502) was higher than the U.S. average in 2006 (489), the only time point to which PISA 2009 performance can be compared in science literacy (figure 8). While U.S. students scored lower than 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.

The PISA 2006 and 2009 OECD averages used in the analysis of trends in science literacy over time are based on the averages of the 34 OECD countries.[12] As a result, the science literacy OECD average score for PISA 2006 differs from previously published reports and is referred to as the OECD trend score.

[12] The OECD excluded the data for Austria from the trend analysis in its report (OECD 2010e) because of a concern over a data collection issue in 2009; however, after consultation with Austrian officials, NCES kept the Austrian data in the U.S. trend reporting.

Figure 8. Average scores of 15-year-old students in the United States and OECD countries on science literacy scale: 2006 and 2009



* *p* < .05. U.S. average is significantly different from the OECD trend score at the .05 level of statistical significance.
***p* < .05. U.S. average in 2006 is significantly different from the U.S. average in 2009 at the .05 level of statistical significance.
NOTE: The PISA science framework was revised in 2006. Because of changes in the framework, it is not possible to compare science learning outcomes from PISA 2000 and 2003 with those from PISA 2006 and 2009. For more details, see OECD (2010e). The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. Scores are reported on a scale from 0 to 1,000. The standard errors of the estimates are shown in table S2 available at http://nces.ed.gov/surveys/pisa/pisa2009tablefigureexhibit.asp.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006 and 2009.

Page intentionally left blank

Further Information

This report provides selected findings from PISA 2009 from a U.S. perspective. Readers who are interested in detailed international findings should consult the OECD PISA 2009 reports (OECD 2010a, 2010b, 2010c, 2010d, 2010e). They may be found at http://www.pisa.oecd.org. PISA data can be analyzed with the PISA Data Explorer, available at http://nces.ed.gov/surveys/international/ide/.

Page intentionally left blank

References

Australian Council for Educational Research (ACER). (2008). *PISA 2009 Main Survey Test Administrator's Manual*. Camberwell, Australia: Author.

Australian Council for Educational Research (ACER). (2009). *PISA 2009 Main Study Data Management Manual, Version 2.* Camberwell, Australia: Author.

Baldi, S., Jin, Y., Skemer, M., Green, P.J., and Herget, D. (2007). *Highlights From PISA 2006: Performance of U.S. 15-Year-Old Students in Science and Mathematics Literacy in an International Context* (NCES 2008–016). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Mislevy, R.J. (1988). Randomization Based Inferences About Latent Variables From Complex Samples. *Psychometrika*, *56*(2): 177-196.

Organization for Economic Cooperation and Development (OECD). (1999). *Measuring Student Knowledge and Skills: A New Framework for Assessment.* Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2004). *Learning for Tomorrow's World - First Results from PISA 2003*. Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2006). *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*. Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006: Science Competencies for Tomorrow's World (Volume I: Analysis.)* Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2009). *PISA 2009 Assessment Framework - Key Competencies in Reading, Mathematics and Science*. Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2010a). *PISA 2009 Results: What Students Know and Can Do - Student Performance in Reading, Mathematics and Science (Volume I).* Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2010b). *PISA 2009 Results: Overcoming Social Background - Equity in Learning Opportunities and Outcomes (Volume II).* Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2010c). *PISA 2009 Results: Learning to Learn - Student Engagement, Strategies and Practices (Volume III).* Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2010d). *PISA 2009 Results: What Makes a School Successful? - Resources, Policies and Practices (Volume IV).* Paris: Author.

Organization for Economic Cooperation and Development (OECD). (2010e). *PISA 2009 Results: Learning Trends - Changes in Student Performance Since 2000 (Volume V).* Paris: Author.

Organization for Economic Cooperation and Development (OECD). (forthcoming). *PISA 2009 Technical Report.* Paris: Author.

Page intentionally left blank

Appendix A: Sample Reading Texts and Items From PISA 2009

After each administration of the Program for International Student Assessment (PISA), the Organization for Economic Cooperation and Development (OECD) releases to the public a subset of items in order to illustrate the content of the assessment. The remaining items are kept secure so they can be used again in a future PISA cycle to measure trends in performance. This appendix contains sample reading texts and items used in the U.S. administration of the PISA 2009 reading assessment. The items illustrate the different aspects of reading assessed by PISA as well as the PISA proficiency levels. The percentage of U.S. students who answered the item correctly is shown, along with the OECD average percentage correct for each item.

Exhibit A-1 shows the PISA 2009 sample items organized by reading aspect and PISA proficiency level. For example, The Play's the Thing question 1 assesses the *integrate and interpret* aspect and is located on the PISA scale at level 6, indicating that it is of high difficulty. The *access and retrieve* aspect and the two lowest proficiency levels (level 1a and level 1b), as well as levels 2, 5 and 6 of *reflect and evaluate* were not covered by the released items on which U.S. students were assessed.

Exhibit A-1. Sample PISA 2009 reading texts and items by reading aspect and PISA proficiency level

	Reading aspect		
Level	Access and retrieve	Integrate and interpret	Reflect and evaluate
Level 6		The Play's the Thing Q1	
Level 5			
Level 4		The Play's the Thing Q3 The Play's the Thing Q4 Cell Phone Safety Q1	Cell Phone Safety Q2
Level 3		Telecommuting Q1 Telecommuting Q3 Cell Phone Safety Q4	Cell Phone Safety Q3 Telecommuting Q2
Level 2		The Play's the Thing Q2	
Level 1a			
Level 1b			

NOTE: The *access and retrieve* aspect and the two lowest proficiency levels (level 1a and level 1b), as well as level 5 of *integrate and interpret* and levels 2, 5 and 6 of *reflect and evaluate* were not covered by the released items on which U.S. students were assessed.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

Exhibit A-2. Example A of PISA 2009 reading assessment: Telecommuting

TELECOMMUTING

The way of the future

Just imagine how wonderful it would be to "telecommute"[1] to work on the electronic highway, with all your work done on a computer or by phone! No longer would you have to jam your body into crowded buses or trains or waste hours and hours travelling to and from work. You could work wherever you want to – just think of all the job opportunities this would open up!

Molly

Disaster in the making

Cutting down on commuting hours and reducing the energy consumption involved is obviously a good idea. But such a goal should be accomplished by improving public transportation or by ensuring that workplaces are located near where people live. The ambitious idea that telecommuting should be part of everyone's way of life will only lead people to become more and more self-absorbed. Do we really want our sense of being part of a community to deteriorate even further?

Richard

[1] "Telecommuting" is a term coined by Jack Nilles in the early 1970s to describe a situation in which workers work on a computer away from a central office (for example, at home) and transmit data and documents to the central office via telephone lines.

Use "Telecommuting" above to answer the questions that follow.

Question 1: TELECOMMUTING

What is the relationship between "The way of the future" and "Disaster in the making"?

A They use different arguments to reach the same general conclusion.

B They are written in the same style but they are about completely different topics.

C They express the same general point of view, but arrive at different conclusions.

(D) They express opposing points of view on the same topic.

Question 2: TELECOMMUTING

What is one kind of work for which it would be difficult to telecommute? Give a reason for your answer.

Plumber. You can't fix someone else's sink from your home! **(full credit)**

Question 3: TELECOMMUTING

Which statement would **both** Molly and Richard agree with?

A People should be allowed to work for as many hours as they want to.

(B) It is not a good idea for people to spend too much time getting to work.

C Telecommuting would not work for everyone.

D Forming social relationships is the most important part of work.

Percentage of students answering correctly

	Level	Aspect		Percentage	*s.e.*
Question 1	Level 3	Integrate and interpret	United States	55	1.6
			OECD average	52	0.2
Question 2	Level 3	Reflect and evaluate	United States	60	1.4
			OECD average	56	0.2
Question 3	Level 3	Integrate and interpret	United States	51	1.6
			OECD average	60	0.2

NOTE: The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. The standard error is noted by *s.e.*
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

Exhibit A-3. Example B of PISA 2009 reading assessment: Cell Phone Safety

CELL PHONE SAFETY

Are cell phones dangerous?

	Yes	No
1.	Radio waves given off by cell phones can heat up body tissue, having damaging effects.	Radio waves are not powerful enough to cause heat damage to the body.
2.	Magnetic fields created by cell phones can affect the way that your body cells work.	The magnetic fields are incredibly weak, and so unlikely to affect cells in our body.
3.	People who make long cell phone calls sometimes complain of fatigue, headaches, and loss of concentration.	These effects have never been observed under laboratory conditions and may be due to other factors in modern lifestyles.
4.	Cell phone users are 2.5 times more likely to develop cancer in areas of the brain adjacent to their phone ears.	Researchers admit it's unclear this increase is linked to using cell phones.
5.	The International Agency for Research on Cancer found a link between childhood cancer and power lines. Like cell phones, power lines also emit radiation.	The radiation produced by power lines is a different kind of radiation, with much more energy than that coming from cell phones.
6.	Radio frequency waves similar to those in cell phones altered the gene expression in nematode worms.	Worms are not humans, so there is no guarantee that our brain cells will react in the same way.

Key Point

Conflicting reports about the health risks of cell phones appeared in the late 1990s.

Key Point

Millions of dollars have now been invested in scientific research to investigate the effects of cell phones.

Key Point

Given the immense numbers of cell phone users, even small adverse effects on health could have major public health implications.

Key Point

In 2000, the Stewart Report (a British report) found no known health problems caused by cell phones, but advised caution, especially among the young, until more research was carried out. A further report in 2004 backed this up.

If you use a cell phone…

Do	**Don't**
Keep the calls short.	Don't use your cell phone when the reception is weak, as the phone needs more power to communicate with the base station, and so the radio-wave emissions are higher.
Carry the cell phone away from your body when it is on standby.	Don't buy a cell phone with a high "SAR" value.[1] This means that it emits more radiation.
Buy a cell phone with a long "talk time." It is more efficient, and has less powerful emissions.	Don't buy protective gadgets unless they have been independently tested.

[1] SAR (specific absorption rate) is a measurement of how much electromagnetic radiation is absorbed by body tissue while using a cell phone.

"Cell Phone Safety" on the previous two pages is from a website.

Use "Cell Phone Safety" to answer the questions that follow.

Question 1: CELL PHONE SAFETY

What is the purpose of the **Key points**?

A To describe the dangers of using cell phones.

(B) To suggest that debate about cell phone safety is ongoing.

C To describe the precautions that people who use cell phones should take.

D To suggest that there are no known health problems caused by cell phones.

Question 2: CELL PHONE SAFETY

"It is difficult to prove that one thing has definitely caused another."

What is the relationship of this piece of information to the Point 4 **Yes** and **No** statements in the table **Are cell phones dangerous?**

A It supports the Yes argument but does not prove it.

B It proves the Yes argument.

(C) It supports the No argument but does not prove it.

D It shows that the No argument is wrong.

Question 3: CELL PHONE SAFETY

Look at Point 3 in the **No** column of the table. In this context, what might one of these "other factors" be? Give a reason for your answer.

Noise – that gives you a headache. **(full credit)**

Question 4: CELL PHONE SAFETY

Look at the table with the heading **If you use a cell phone…**

Which of these ideas is the table based on?

A There is no danger involved in using cell phones.

B There is a proven risk involved in using cell phones.

(C) There may or may not be danger involved in using cell phones, but it is worth taking precautions.

D There may or may not be danger involved in using cell phones, but they should not be used until we know for sure.

E The **Do** instructions are for those who take the threat seriously, and the **Don't** instructions are for everyone else.

Percentage of students answering correctly

	Level	Aspect		Percentage	*s.e.*
Question 1	Level 4	Integrate and interpret	United States	48	1.5
			OECD average	45	0.2
Question 2	Level 4	Reflect and evaluate	United States	42	1.6
			OECD average	35	0.2
Question 3	Level 3	Reflect and evaluate	United States	52	1.7
			OECD average	54	0.3
Question 4	Level 3	Integrate and interpret	United States	68	1.4
			OECD average	62	0.2

NOTE: The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. The standard error is noted by *s.e.*
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

Exhibit A-4. Example C of PISA 2009 reading assessment: The Play's the Thing

THE PLAY'S THE THING

Takes place in a castle by the beach in Italy.

FIRST ACT

Ornate guest room in a very nice beachside castle. Doors on the right and left. Sitting room set in the middle of the stage: couch, table, and two armchairs. Large windows at the back. Starry night. It is dark on the stage. When the curtain goes up we hear men conversing loudly behind the door on the left. The door opens and three tuxedoed gentlemen enter. One turns the light on immediately. They walk to the center in silence and stand around the table. They sit down together, Gál in the armchair to the left, Turai in the one on the right, Ádám on the couch in the middle. Very long, almost awkward silence. Comfortable stretches. Silence. Then:

GÁL

Why are you so deep in thought?

TURAI

I'm thinking about how difficult it is to begin a play. To introduce all the principal characters in the beginning, when it all starts.

ÁDÁM

I suppose it must be hard.

TURAI

It is – devilishly hard. The play starts. The audience goes quiet. The actors enter the stage and the torment begins. It's an eternity, sometimes as much as a quarter of an hour before the audience finds out who's who and what they are all up to.

GÁL

Quite a peculiar brain you've got. Can't you forget your profession for a single minute?

TURAI

That cannot be done.

GÁL

Not half an hour passes without you discussing theater, actors, plays. There are other things in this world.

TURAI

There aren't. I am a dramatist. That is my curse.

GÁL

You shouldn't become such a slave to your profession.

TURAI

If you do not master it, you are its slave. There is no middle ground. Trust me, it's no joke starting a play well. It is one of the toughest problems of stage mechanics. Introducing your characters promptly. Let's look at this scene here, the three of us. Three gentlemen in tuxedoes. Say they enter not this room in this lordly castle, but rather a stage, just when a play begins. They would have to chat about a whole lot of uninteresting topics until it came out who we are. Wouldn't it be much easier to start all this by standing up and introducing ourselves? *Stands up.* Good evening. The three of us are guests in this castle. We have just arrived from the dining room where we had an excellent dinner and drank two bottles of champagne. My name is Sándor Turai, I'm a playwright, I've been writing plays for thirty years, that's my profession. Full stop. Your turn.

GÁL

Stands up. My name is Gál, I'm also a playwright. I write plays as well, all of them in the company of this gentleman here. We are a famous playwright duo. All playbills of good comedies and operettas read: written by Gál and Turai. Naturally, this is my profession as well.

GÁL and TURAI

Together. And this young man …

ÁDÁM

Stands up. This young man is, if you allow me, Albert Ádám, twenty-five years old, composer. I wrote the music for these kind gentlemen for their latest operetta. This is my first work for the stage. These two elderly angels have discovered me and now, with their help, I'd like to become famous. They got me invited to this castle. They got my dress-coat and tuxedo made. In other words, I am poor and unknown, for now. Other than that I'm an orphan and my grandmother raised me. My grandmother has passed away. I am all alone in this world. I have no name, I have no money.

TURAI

But you are young.

GÁL

And gifted.

ÁDÁM

And I am in love with the soloist.

TURAI

You shouldn't have added that. Everyone in the audience would figure that out anyway.

They all sit down.

TURAI

Now wouldn't this be the easiest way to start a play?

GÁL

If we were allowed to do this, it would be easy to write plays.

TURAI

Trust me, it's not that hard. Just think of this whole thing as …

GÁL

All right, all right, all right, just don't start talking about the theater again. I'm fed up with it. We'll talk tomorrow, if you wish.

"The Play's the Thing" is the beginning of a play by the Hungarian dramatist Ferenc Molnár.

Use "The Play's the Thing" on the previous two pages to answer the questions that follow. (Note that line numbers are given in the margin of the script to help you find parts that are referred to in the questions.)

Question 1: THE PLAY'S THE THING

What were the characters in the play doing **just before** the curtain went up?

Had dinner and drank. **(full credit)**

Question 2: THE PLAY'S THE THING

"It's an eternity, sometimes as much as a quarter of an hour… " (lines 29-30)

According to Turai, why is a quarter of an hour "an eternity"?

A It is a long time to expect an audience to sit still in a crowded theater.

(B) It seems to take forever for the situation to be clarified at the beginning of a play.

C It always seems to take a long time for a dramatist to write the beginning of a play.

D It seems that time moves slowly when a significant event is happening in a play.

Question 3: THE PLAY'S THE THING

A reader said, "Ádám is probably the most excited of the three characters about staying at the castle."

What could the reader say to support this opinion? Use the text to give a reason for your answer.

He must be happy to be with the two guys who can make him famous. **(full credit)**

Question 4: THE PLAY'S THE THING

Overall, what is the dramatist Molnár doing in this extract?

A He is showing the way that each character will solve his own problems.

B He is making his characters demonstrate what an eternity in a play is like.

C He is giving an example of a typical and traditional opening scene for a play.

(D) He is using the characters to act out one of his own creative problems.

Percentage of students answering correctly

	Level	Aspect		Percentage	*s.e.*
Question 1	Level 6	Integrate and interpret	United States	13	1.0
			OECD average	13	0.2
Question 2	Level 2	Integrate and interpret	United States	61	1.2
			OECD average	66	0.2
Question 3	Level 4	Integrate and interpret	United States	54	1.7
			OECD average	49	0.3
Question 4	Level 4	Integrate and interpret	United States	44	1.6
			OECD average	46	0.2

NOTE: The Organization for Economic Cooperation and Development (OECD) average is the average of the national averages of the OECD member countries, with each country weighted equally. The standard error is noted by *s.e.*
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2009.

Page intentionally left blank

Appendix B: Technical Notes

The Program for International Student Assessment (PISA) is an international assessment that measures 15-year-olds' performance in reading literacy, mathematics literacy, and science literacy. First implemented in 2000, PISA is coordinated by the Organization for Economic Cooperation and Development (OECD), an intergovernmental organization of 34 member countries. In the fourth cycle (PISA 2009), reading literacy was the major focus. This appendix describes features of the PISA 2009 methodology, including sample design, test design, scoring, data reliability, and analysis variables. For further details about the assessment and any of the topics discussed here, see the OECD's *PISA 2009 Technical Report (forthcoming).*

International Requirements for Sampling, Data Collection, and Response Rates

To provide valid estimates of student achievement and characteristics, the sample of PISA students had to be selected in a way that represented the full population of 15-year-old students in each country. The international desired population in each country consisted of 15-year-olds attending both publicly and privately controlled schools in grade 7 and higher. A minimum of 4,500 students from a minimum of 150 schools was required in each country. The international guidelines specified that within schools, a sample of 35 students was to be selected in an equal probability sample unless fewer than 35 students age 15 were available (in which case all students were selected). International standards required that students in the sample be 15 years and 3 months to 16 years and 2 months at the beginning of the testing period. In the United States, sampled students were born between July 1, 1993, and June 30, 1994. The international standard for the maximum length of the testing period was 42 days, but the United States requested and was granted permission to expand the testing window to 60 days (from September 21, 2009, to November 19, 2009) in order to accommodate school requests.[1] Each country collected its own data, following international guidelines and specifications.

[1] Most countries conducted testing from March through August of 2009. The United States and the United Kingdom were given permission to move the testing dates to September through November in an effort to improve response rates. The range of eligible birthdates was adjusted so that the mean age remained the same (i.e., 15 years and 3 months to 16 years and 2 months at the beginning of the testing period). In 2003, the United States conducted PISA in the spring and fall and found no significant difference in student performance between the two time points.

The school response-rate target was 85 percent for all countries. A minimum of 65 percent of schools from the original sample of schools was required to participate for a country's data to be included in the international database. Countries were allowed to use replacement schools (selected during the sampling process) to increase the response rate once the 65 percent benchmark had been reached.

PISA 2009 also required a minimum participation rate of 80 percent of sampled students from schools within each country. A student was considered to be a participant if he or she participated in the first testing session or a follow-up or makeup testing session. Data from countries not meeting this requirement could be excluded from international reports.

PISA's intent was to be as inclusive as possible. The guidelines allowed schools to be excluded for approved reasons (for example, schools in remote regions, very small schools, or special education schools could be excluded). Schools used the following international guidelines on student exclusions:

- **Students with functional disabilities.** These were students with a moderate to severe permanent physical disability such that they cannot perform in the PISA testing environment.
- **Students with intellectual disabilities.** These were students with a mental or emotional disability and who have been tested as cognitively delayed or who are considered in the professional opinion of qualified staff to be cognitively delayed such that they cannot perform in the PISA testing environment.
- **Students with insufficient language experience.** These were students who meet the three criteria of not being native speakers in the assessment language, having limited proficiency in the assessment language, and having less than one year of instruction in the assessment language.

Overall estimated exclusions (including both school and student exclusions) were to be under 5 percent of the PISA target population.

Quality monitors from the PISA Consortium visited a sample of schools in every country to ensure that testing procedures were conducted in a consistent manner.

Sampling, Data Collection, and Response Rates in the United States

The PISA 2009 school sample was drawn for the United States in July 2008 by the international PISA Consortium. The U.S. sample for 2009 was drawn using a two-stage sampling process. The first stage was a sample of schools and the second stage was a sample of students within schools. The sample design for PISA 2009 was a stratified systematic sample, with sampling probabilities proportional to the estimated number of 15-year-old students in the school based on grade enrollments. The PISA sample was stratified into eight explicit groups based on control of school (public or private) and region of the country (Northeast, Central, West, Southeast).[2] Within each stratum, the frame was implicitly stratified (i.e., sorted for sampling) by five categorical stratification variables: grade range of the school (five categories); type of location relative to populous areas (city, suburb, town, rural);[3] first three digits of the zip code; combined percentage of Black, Hispanic, Asian, Pacific Islander, and American Indian/Alaska Native students (above or below 15 percent); and estimated enrollment of 15-year-olds. The sampling employed techniques to minimize overlap with the High School Longitudinal Study of 2009 (which was collecting data in the same school year) and to undersample very small schools (those with an estimate of fewer than twenty-one 15-year-old students).

Following the PISA guidelines, at the same time as the PISA sample was selected, replacement schools were identified by assigning the two schools neighboring the sampled school in the frame as replacements. There were several constraints on the assignment of substitutes. One sampled school was not allowed to substitute for another, and a given school could not be assigned to substitute for more than one sampled school. Furthermore, substitutes were required to be in the same explicit stratum as the sampled school. If the sampled school was the first or last school in the stratum, the second school following or preceding the sampled school was identified as the substitute. One school was designated a first replacement and the other a second replacement. If an original school refused to participate, the first replacement was then contacted. If that school also refused to participate, the second school was contacted.

The U.S. PISA 2009 school sample consisted of 236 schools. This number was increased from the international minimum requirement of 150 to offset school nonresponse and reduce design effects. Schools were selected with probability proportionate to the school's estimated enrollment of 15-year-olds. The data for public schools were from the 2005–06 Common Core of Data (CCD), and the data for private schools were from the 2005–06 Private School Universe Survey (PSS). Any school containing at least one 7th- through 12th-grade class in school year 2005–06 was included in the school sampling frame. Participating schools provided a list of 15-year-old students (typically in August or September 2009), and a sample of 42 students was selected within each school in an equal probability sample. The overall sample design for the United States was intended to approximate a self-weighting sample of students as much as possible, with each 15-year-old student having an equal probability of being selected.

In the United States, for a variety of reasons reported by school administrators (such as increased testing requirements at the national, state, and local levels; concerns about the timing of the PISA assessment; and loss of learning time), many schools in the original sample declined to participate. Of the 236 original sampled schools, 208 were eligible (22 schools did not have any 15-year-olds enrolled, 5 had closed, and 1 was ineligible because all of its students were also enrolled in other "home" schools), and 145 agreed to participate. The weighted school response rate before replacement was 68 percent, requiring the United States to conduct a nonresponse bias analysis, which was used by the PISA Consortium and the OECD to evaluate the quality of the final sample.[4] In addition to the 145 participating original schools, 20 replacement schools also participated, for a total of 165 participating schools (or a 78 percent overall school response rate).[5]

[2] The Northeast region consists of Connecticut, Delaware, the District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. The Central region consists of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, Wisconsin, and South Dakota. The West region consists of Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oklahoma, Oregon, Texas, Utah, Washington, and Wyoming. The Southeast region consists of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia.

[3] These types are defined as follows: (1) "city" is territory inside an urbanized area with a core population of 50,000 or more and inside a principal city; (2) "suburb" is territory inside an urbanized area with a core population of 50,000 or more and outside a principal city; (3) "town" is territory inside an urban cluster with a core population between 25,000 and 50,000; and (4) "rural" is territory not in an urbanized area or urban cluster.

[4] NCES requires a nonresponse bias analysis for any survey with a weighted response rate below 85 percent. OECD requires a nonresponse bias analysis from countries with weighted school response rates between 65 and 85 percent.

[5] Response rates reported here are based on the formula used in the international report and are not consistent with NCES standards. A more conservative way to calculate the response rate would be to include replacement schools that participated in the denominator as well as the numerator, and to add replacement schools that were hard refusals to the denominator. This results in a weighted school response rate of 64 percent.

A total of 6,677 students were sampled for the assessment. Of these students, 273 were deemed ineligible because they had left the school between the samping and assessment date. Of the eligible 6,404 sampled students, an additional 339 were excluded using the decision criteria described earlier, for a weighted exclusion rate of 5 percent at the student level.

Of the 6,065 remaining sampled students, a total of 5,233 participated in the assessment in the United States for an overall weighted student response rate of 87 percent.

A bias analysis was conducted in the United States to address potential problems in the data owing to school nonresponse. To compare PISA participating schools and nonparticipating schools, it was necessary to match the sample of schools back to the sample frame to detect as many characteristics as possible that might provide information about the presence of nonresponse bias. Frame characteristics were taken from the 2005–06 CCD for public schools and from the 2005–06 PSS for private schools. The available school characteristics included affiliation (public or private), community type, region, number of age-eligible students, total number of students, and percentage of various racial/ethnic groups (Asian or Pacific Islander, non-Hispanic; Black, non-Hispanic; Hispanic; American Indian or Alaska Native, non-Hispanic; and White, non-Hispanic). The percentage of students eligible for free or reduced-price lunch was available for public schools only.

Comparing frame characteristics for participating schools and nonparticipating schools is not always a good measure of nonresponse bias if the characteristics are unrelated or weakly related to more substantive items in the survey; however, this was the only approach available given that no comparable school- or student-level achievement data were available.

For categorical variables, the hypothesis of independence between the characteristics and response status was tested using a chi-square statistic. For continuous variables, summary means were calculated and compared using t tests. In addition to these tests, logistic regression models were employed to identify whether any of the frame characteristics were significant in predicting response status. All analyses were performed using WesVar, a statistical software package. The school base weights used in these analyses did not include a nonresponse adjustment factor. The base weight for each original school was calculated as the reciprocal of the probability of selection times the number of eligible students in the school. The base weight for each replacement school was set equal to the base weight of the original school it replaced.

The only variable for which there were statistically significant differences between participating schools and all sampled schools was the percentage of students at the school eligible for free or reduced-price lunch ($t = 2.30$, $p = .02$). On average, participating schools had a higher percentage of students from lower income families (mean = 35.4 percent, *s.e.*= 1.95) who were eligible for free or reduced-price lunch than did all sampled schools (mean = 34.1 percent, *s.e.*= 1.70).

Test Development

The 2009 assessment instruments were developed by international experts and PISA Consortium test developers, and items were reviewed by representatives of each country for possible bias and relevance to PISA's goals. The assessment included items submitted by participating countries as well as items that were developed by the Consortium's test developers.

The final assessment consisted of 102 reading items, 36 mathematics items, and 52 science items allocated to 13 test booklets. Each booklet was made up of 4 test clusters. Altogether there were 7 reading clusters, 3 mathematics clusters, and 3 science clusters. The clusters were allocated in a rotated design to the 13 booklets. The average number of items per cluster was 15 items for reading, 12 items for mathematics, and 17 items for science. Each cluster was designed to average 30 minutes of test material. Each student took one booklet, with about 2 hours worth of testing material. Approximately half of the items were multiple-choice, about 20 percent were closed or short response types (for which students wrote an answer that was simply either correct or incorrect), and about 30 percent were open constructed responses (for which students wrote answers that were graded by trained scorers using an international scoring guide). In PISA 2009, every student answered reading items. Not all students answered mathematics and/or science items.

In addition to the cognitive assessment, students also received a 30-minute questionnaire designed to provide information about their backgrounds, attitudes, and experiences in school. Principals in schools where PISA was administered also received a 30-minute questionnaire about their schools.

Translation and Adaptation

Source versions of all instruments (assessment booklets, questionnaires, and manuals) were prepared in English and French and translated into the primary language or languages of instruction in each country. PISA recommended that countries prepare and consolidate independent translations from both source versions and provided precise translation guidelines that included a description of the features each item was measuring and statistical analysis from the field trial. In cases for which one source language was used, independent translations were required and discrepancies reconciled. In addition, it was sometimes necessary to adapt the instrument for cultural purposes, even in nations such as the United States that use English as the primary language of instruction. For example, words such as "lift" might be adapted to "elevator" for the United States. The PISA Consortium verified the national adaptation of all instruments. Electronic copies of printed materials were sent to the PISA Consortium for a final visual check prior to data collection.

Test Administration and Quality Assurance

PISA 2009 emphasized the use of standardized procedures in all countries. Each country collected its own data, based on a manual provided by the PISA Consortium (ACER 2008) to explain the survey's implementation, including precise instructions for the work of school coordinators and scripts for test administrators to use in testing sessions. Test administration in the United States was coordinated by professional staff trained according to the international guidelines. School staff members were asked to assist only with listing students, identifying space for testing in the school, and specifying any parental consent procedures needed for sampled students. Students were allowed to use calculators, and U.S. students were provided calculators; however, no information on the availability of calculators was collected internationally.

At some schools, the PISA assessment was administered to students outside of normal school hours to address schools' concerns about the potential negative effect on students of the loss of instructional time. In the United States, tests were administered during normal school hours at 155 schools (94 percent), outside of normal school hours at 4 schools (2 percent), and on Saturdays at 6 schools (4 percent).

Test administrations were observed in a sample of schools in each country by a PISA Quality Monitor (PQM) who was engaged by the PISA Consortium. The sample schools were selected jointly by the PISA Consortium and the PQM. In the United States, 7 schools were observed by the PQM. The PQM's primary responsibility was to document the extent to which testing procedures in schools were implemented in accordance with test administration procedures. The PQM's observations in U.S. schools indicated that international procedures for data collection were applied consistently.

Scoring

A significant proportion of the PISA assessment was devoted to items requiring constructed responses. The scoring of these responses was the responsibility of each country. The process of scoring these items was an important step in ensuring the quality and comparability of the PISA data.

The PISA Consortium developed detailed scoring guides, scoring training materials, and scorer recruitment materials and led international training sessions on scoring. Those who attended the international training on scoring then led the training of national scoring teams.

For each test item, the scoring guide described the intent of the question and how to score the students' responses. This description included the credit labels—full credit, partial credit, or no credit—attached to the possible categories of response. In addition, the scoring guides included real examples of students' responses accompanied by a rationale for their classification for purposes of clarity and illustration.

To examine the consistency of this marking process in more detail within each country and to estimate the magnitude of the variance components associated with the use of scorers, the PISA Consortium conducted an interscorer reliability study on a subsample of assessment booklets. Homogeneity analysis was applied to the national sets of multiple scoring and compared with the results of the field trial. A full description of this process and the results can be found in the OECD's *PISA 2009 Technical Report (forthcoming)*.

Data Entry and Cleaning

Data entry was the responsibility of each country. The data collected for PISA 2009 were entered into data files with a common international format, as specified in the *PISA 2009 Main Study Data Management Manual, Version 2* (ACER 2009). Data entry was completed using specialized software that allowed data to be merged into KeyQuest, a common data processing software application developed by Australian Council for Educational Research (ACER) for use by participating countries. The software facilitated the checking and correction of data by providing various data consistency checks. The data were then sent to ACER for cleaning. ACER's role at this point was to check that the international data structure was followed, check the identification system within and between files, correct single case problems manually, and apply standard cleaning procedures to questionnaire files. Results of the data cleaning process were documented and shared with the national project managers and included specific questions when required. The national project manager then provided ACER with revisions to coding or solutions for anomalies. ACER then compiled background univariate statistics and preliminary classical and Rasch item analysis. Detailed information on the entire data entry and cleaning process can be found in the OECD's *PISA 2009 Technical Report (forthcoming)*.

Weighting

The use of sampling weights is necessary for the computation of statistically sound, nationally representative estimates. Adjusted survey weights adjust for the probabilities of selection for individual schools and students, for school or student nonresponse, or for errors in estimating the size of the school or the number of 15-year-olds in the school at the time of sampling. Survey weighting for all countries and other education systems participating in PISA 2009 was coordinated by Westat, as part of the PISA Consortium.

The school base weight was defined as the reciprocal of the school's probability of selection times the number of eligible students in the school. (For replacement schools, the school base weight was set equal to the original school it replaced.) The student base weight was given as the reciprocal of the probability of selection for each selected student from within a school.

The product of these base weights was then adjusted for school and student nonresponse. The school nonresponse adjustment was done individually for each country using the explicit strata defined as part of the sample design. In the case of the United States, two variables were used: school control and census region. The student nonresponse adjustment was done within cells based first on their school nonresponse rate and their explicit stratum; within that, grade and sex were used when possible. Grade and sex were collected for students in all countries on the student tracking form. All PISA analyses were conducted using these adjusted sampling weights. For more information on the nonresponse adjustments, see the OECD's *PISA 2009 Technical Report (forthcoming)*.

Scaling of Student Test Data

Thirteen versions of the PISA test booklet were created, each containing a different subset of items. The fact that each student completed only a subset of items means that classical test scores, such as the percent correct, are not accurate measures of student performance. Instead, scaling techniques were used to establish a common scale for all students. For PISA 2009, item response theory (IRT) was used to estimate average scores for reading, mathematics, and science literacy for each country, as well as for three reading literacy subscales: *accessing and retrieving information, integrating and interpreting,* and *reflecting and evaluating.* [6]

IRT identifies patterns of response and uses statistical models to predict the probability of answering an item correctly as a function of the students' proficiency in answering other questions. With this method, the performance of a sample of students in a subject area or subarea can be summarized on a simple scale or series of scales, even when students are administered different items.

Scores for students are estimated as plausible values because each student completed only a subset of items. Five plausible values were estimated for each student for each scale. These values represent the distribution of potential scores for all students in the population with similar characteristics and identical patterns of item response. Statistics describing performance on the PISA reading, mathematics, and science literacy scales are based on plausible values.[7]

Proficiency Levels

In addition to a range of scale scores as the basic form of measurement, PISA describes student proficiency in terms of levels. Higher levels represent the knowledge, skills, and capabilities needed to perform tasks of increasing complexity. As a result, the findings are reported in terms of percentages of the student population at each of the predefined levels.

To determine the performance levels and cut scores on the literacy scales, IRT techniques were used. With IRT techniques, it is possible to simultaneously estimate the ability of all students taking the PISA assessment, as well as the difficulty of all PISA items. Then estimates of student ability and item difficulty can be mapped

[6] The combined reading literacy scale is made up of all items in the three subscales. However, the combined reading scale and the three subscales are each computed separately through IRT models. Therefore, the combined reading scale score is not the average of the three subscale scores.

[7] For theoretical and empirical justification of the procedures employed, see Mislevy (1988).

on a single continuum. The relative ability of students taking a particular test can be estimated by considering the percentage of test items they get correct. The relative difficulty of items in a test can be estimated by considering the percentage of students getting each item correct. In PISA, all students within a level are expected to answer at least half of the items from that level correctly. Students at the bottom of a level are able to provide the correct answers to about 52 percent of all items from that level, have a 62 percent chance of success on the easiest items from that level, and have a 42 percent chance of success on the hardest items from that level. Students in the middle of a level have a 62 percent chance of correctly answering items of average difficulty for that level (an overall response probability of 62 percent). Students at the top of a level are able to provide the correct answers to about 70 percent of all items from that level, have a 78 percent chance of success on the easiest items from that level, and have a 62 percent chance of success on the hardest items from that level. Students just below the top of a level would score less than 50 percent on an assessment at the next higher level. Students at a particular level demonstrate not only the knowledge and skills associated with that level but also the proficiencies defined by lower levels. Patterns of responses for students below level 1b for reading literacy and below level 1 for mathematics and science literacy suggest that these students are unable to answer at least half of the items from those levels correctly. For details about the approach to defining and describing the PISA levels and establishing the cut scores, see the OECD's *PISA 2009 Technical Report (forthcoming)*.

The reading proficiency level ranges are below level 1b (a score less than or equal to 262.04); level 1b (a score greater than 262.04 and less than or equal to 334.75); level 1a (a score greater than 334.75 and less than or equal to 407.47); level 2 (a score greater than 407.47 and less than or equal to 480.18); level 3 (a score greater than 480.18 and less than or equal to 552.89); level 4 (a score greater than 552.89 and less than or equal to 625.61); level 5 (a score greater than 625.61 and less than or equal to 698.32); and level 6 (a score greater than 698.32). The math profiency level ranges are below level 1 (a score less than or equal to 357.77); level 1 (a score greater than 357.77 and less than or equal to 420.07); level 2 (a score greater than 420.07 and less than or equal to 482.38); level 3 (a score greater than 482.38 and less than or equal to 544.68); level 4 (a score greater than 544.68 and less than or equal to 606.99); level 5 (a score greater than 606.99 and less than or equal to 669.30); and level 6 (a score greater than 669.30). Science proficiency level ranges are below level 1 (a score less than or equal to 334.94); level 1 (a score greater than 334.94 and less than or equal to 409.54); level 2 (a score greater than 409.54 and less than or equal to 484.14); level 3 (a score greater than 484.14 and less than or equal to 558.73); level 4 (a score greater than 558.73 and less than or equal to 633.33); level 5 (a score greater than 633.33 and less than or equal to 707.93); and level 6 (a score greater than 707.93).

Data Limitations

As with any study, there are limitations to PISA 2009 that should be taken into consideration. Estimates produced using data from PISA 2009 are subject to two types of error: nonsampling and sampling errors. Nonsampling errors can be due to errors made in the collection and processing of data. Sampling errors can occur because the data were collected from a sample rather than a complete census of the population.

Nonsampling Errors

"Nonsampling error" is a term used to describe variations in the estimates that may be caused by population coverage limitations, nonresponse bias, and measurement error, as well as data collection, processing, and reporting procedures. For example, the sampling frame was limited to regular public and private schools in the 50 states and the District of Columbia and cannot be used to represent Puerto Rico or other jurisdictions. The sources of nonsampling errors are typically problems such as unit and item nonresponse, the differences in respondents' interpretations of the meaning of survey questions, and mistakes in data preparation. Some of these issues (particularly school nonresponse) are discussed earlier in the section on U.S. sampling and data collection.

There are four kinds of missing data at the item level. "Nonresponse" data occur when a respondent is expected to answer an item but no response is given. Responses that are "missing or invalid" occur in multiple-choice items for which an invalid response is given. (The missing or invalid code is not used for open-ended questions.) An item is "not applicable" when it is not possible for the respondent to answer the question. Finally, items that are "not reached" are consecutive missing values starting from the end of each test session. All four kinds of missing data are coded differently in the PISA 2009 database.

Sampling Errors

Sampling errors occur when a discrepancy between a population characteristic and the sample estimate arises because not all members of the target population are sampled for the survey. The size of the sample relative to the population and the variability of the population characteristics both influence the magnitude of sampling error. The particular sample of 15-year-old students from fall 2009 was just one of many possible samples that could have been selected. Therefore, estimates produced from the PISA 2009 sample may differ from estimates that would have been produced had another sample of students been selected. This type of variability is called sampling error because it arises from using a sample of 15-year-old students in 2009 rather than all 15-year-old students in that year.

One potential source of sampling error for PISA 2009 is that the weight for a replacement school was based on the weight for the school originally selected. These schools were typically very similar in size and other characteristics (the replacement schools were adjacent to the original school on the sorted list of schools); however, there could be some error associated with this method. A second potential source of sampling error could occur if the enrollment lists used for sampling were not up to date.

The standard error is a measure of the variability owing to sampling when estimating a statistic. The approach used for calculating sampling variances in PISA was the Fay method of Balanced Repeated Replication (BRR). This method of producing standard errors uses information about the sample design to produce more accurate standard errors than would be produced using simple random sample assumptions. Thus, the standard errors that are reported here can be used as a measure of the precision expected from this particular sample.

In keeping with NCES standards, 95 percent confidence intervals are used for this report. Thus, there is a 95 percent chance that the true average in the population falls within the range of 1.96 times the standard error above or below the estimated score.

Descriptions of Background Variables

In this report, PISA 2009 results are provided for groups of students with different demographic characteristics. Definitions of subpopulations are as follows:

Sex: Results are reported separately for male students and female students.

Race/ethnicity: In the United States, students' race/ethnicity was obtained through student responses to a two-part question in the student questionnaire. Students were asked first whether they were Hispanic or Latino and then whether they were members of the following racial groups: White, Black, Asian, American Indian or Alaska Native, or Native Hawaiian/Other Pacific Islander. Multiple responses to the race classification question were allowed. Results are shown separately for White (non-Hispanic) students, Black (non-Hispanic) students, Hispanic students, Asian (non-Hispanic) students, American Indian or Alaska Native (non-Hispanic) students, Native Hawaiian/Other Pacific Islander (non-Hispanic) students, and non-Hispanic students who selected two or more races.

Socioeconomic levels of families served by school: In the United States, an indicator of socioeconomic level of families in public schools was obtained from respondents (principals or their designees) to the school questionnaire; the respondents were asked to report the percentage of students at the school in the 2008–2009 school year who were eligible to receive free or reduced-price lunch through the National School Lunch Program. The answers were grouped into five categories: less than 10 percent; 10 to 24.9 percent; 25 to 49.9 percent; 50 to 74.9 percent; and 75 percent or more. Analysis was limited to public schools. Missing data on this variable were replaced with measures taken from the CCD.

Confidentiality and Disclosure Limitations

The PISA 2009 data are hierarchical and include school and student data from the participating schools. Confidentiality analyses for the United States were designed to provide reasonable assurance that public-use data files issued by the PISA Consortium would not allow identification of individual U.S. schools or students when compared against other public-use data collections. Disclosure limitations included identifying and masking potential disclosure risk to PISA schools and including an additional measure of uncertainty to school and student identification through random swapping of data elements within the student and school files.

Statistical Procedures

Comparisons made in the text of this report have been tested for statistical significance. For example, in the commonly made comparison of OECD averages to U.S. averages, tests of statistical significance were used to establish whether or not the observed differences from the U.S. average were statistically significant.

The estimation of the standard errors that are required to undertake the tests of significance is complicated by the complex sample and assessment designs, both of which generate error variance. Together they mandate a set of statistically complex procedures for estimating the correct standard errors. As a consequence, the estimated standard errors contain a sampling variance component estimated by BRR. Where the assessments are concerned, there is an additional imputation variance component arising from the assessment design. Details on the BRR procedures used can be found in the *PISA 2009 Technical Report (forthcoming)*.

In almost all instances, the tests for significance used were standard *t* tests. These fell into two categories according to the nature of the comparison being made: comparisons of independent samples and comparisons of nonindependent samples. In PISA, country groups are independent.

In simple comparisons of independent averages, such as the average score of country 1 with that of country 2, the following formula was used to compute the *t* statistic:

$$t = (est_1 - est_2) \ / \ \text{SQRT} \ [(se_1)^2 + (se_2)^2],$$

where est_1 and est_2 are the estimates being compared (e.g., averages of country 1 and country 2) and se_1 and se_2 are the corresponding standard errors of these averages.

The second type of comparison used in this report occurred when comparing differences of nonsubset, nonindependent groups. When this occurs, the correlation and related covariance between the groups must be taken into account (for example, when comparing the average scores of males and females within the United States).

How are scores—such as those for males and females—correlated? Suppose that in the school sample, a coeducational school attended by low achievers is replaced by a coeducational school attended by high achievers. The country mean will increase slightly, as well as the means for males and females. If such a school replacement process is continued, the average scores of males and the average scores of females will likely increase in a similar pattern. Indeed, a coeducational school attended by high-achieving males is usually also attended by high-achieving females. Therefore, the covariance between the males' scores and the females' scores is likely to be positive.

To determine whether the performance of females differs from the performance of males, the standard error of the difference that takes into account the covariance between females' scores and males' scores needs to be estimated. The estimation of the covariance requires the selection of several samples and then the analysis of the variation of males' means in conjunction with females' means. Such a procedure is, of course, unrealistic. Therefore, as for any computation of a standard error in PISA, replication methods using the supplied replicate weights were used to estimate the standard error of a difference. Use of the replicate weights implicitly incorporates the covariance between the two estimates into the estimate of the standard error of the difference.

To test such comparisons, the following formula was used to compute the *t* statistic:

$$t = (est_{grp1} - est_{grp2}) \ / se \ (est_{grp1} - est_{grp2}),$$

where est_{grp1} and est_{grp2} are the nonindependent group estimates being compared and $se\ (est_{grp1} - est_{grp2})$ is the standard error of the difference calculated using BRR to account for any covariance between the estimates for the two nonindependent groups.

Appendix C: U.S. PISA 2009 Steering Committee

Bruce Alberts
University of California, San Francisco, CA

Alicia Alonzo
Michigan State University, East Lansing, MI

Richard Askey
University of Wisconsin, Madison, WI

Gina Biancarosa
Stanford University, Stanford, CA

Jack Buckley (former member)
New York University, New York, NY

Carol Jago
California Reading & Literature Project, Pacific Palisades, CA

Michael Kamil
Stanford University, Stanford, CA

Tom Loveless
The Brookings Institution, Washington, DC

Elizabeth Birr Moje
University of Michigan, Ann Arbor, MI

Joan Ferrini-Mundy (ex-officio member)
National Science Foundation, Arlington, VA

Ramsay Selden
Consultant, Alexandria, VA

Mark Wilson
University of California, Berkeley, CA

Page intentionally left blank

Appendix D: Comparing PISA and NAEP

In the United States, nationally representative data on student achievement come primarily from two sources: the National Assessment of Educational Progress (NAEP)—also known as the "Nation's Report Card"—and the United States' participation in international assessments, including the Program for International Student Assessment (PISA). While PISA may appear to have significant similarities with NAEP, each was designed to serve a different purpose, assesses different target populations, and are based on separate and unique frameworks and items. As such, PISA and NAEP provide different, and complementary, information about student performance.

As reading was the major domain assessed in PISA 2009, NCES sought to compare the content assessed by PISA and NAEP 2009 assessments. It convened an external panel of reading experts to examine the PISA assessment in relation to the NAEP assessment at grades 8 and 12. The group examined and compared reading frameworks, passages, and items between the international and national assessments, looking at the following: how each assessment defined reading; how the domain was organized in the frameworks; the nature, length, and difficulty of the reading passages; and the cognitive processes in which students were asked to engage. This section highlights some of the main findings; additional details on the comparison study will be included in a technical report to be released with the U.S. national PISA dataset at a later date.

- The PISA and NAEP definitions of reading both identify reading as a constructive process that involves interaction between the reader and the text and both focus on understanding and using written text. There are subtle differences, however. PISA's definition emphasizes the use of reading for personally-defined goals and growth and for participation in society. NAEP's definition reflects the notion that readers draw on the ideas and information they acquire from text to meet a particular purpose or situational need.
- There are some similarities in how the frameworks are organized—both NAEP and PISA specify a cognitive dimension and a range of text types. However, PISA includes some organizational elements that NAEP does not and there are differences in how the cognitive categories are defined and in the text types targeted for inclusion. For example, PISA aims to include more noncontinuous texts than NAEP does.
- Individual reading passages in PISA are shorter on average than those used in the NAEP grade 8 and grade 12 assessments. Students are asked an average of 3.6 items per reading passage on PISA but an average of about 10 items per passage on the NAEP grade 8 and 12 assessments. Based on readability analyses, PISA passages are on average more difficult than the NAEP eighth-grade passages and similar to NAEP twelfth-grade passages.
- The panel also considered whether the PISA and NAEP passages, in terms of text type and format, could be found on the other assessment, based on how the respective frameworks described the intended texts; this is referred to as the "fit" of passages to a framework. The panel found that PISA passages[1] tended to fit better to the NAEP framework than did the NAEP passages to the PISA framework, though a substantial number of passages from both assessments were deemed not interchangeable. About half the NAEP eighth-grade and two-thirds of the NAEP twelfth-grade passages were considered to not fit within the PISA framework and about two-fifths of PISA passages were considered to not fit within the NAEP framework.
- PISA and NAEP passages differ with respect to "authenticity." The NAEP framework emphasizes the authenticity of text and notes a commitment to selecting high-quality, authentic stimulus materials that students are likely to encounter both in school and out of school. There is some flexibility in excerpting stimulus material, but texts are not edited prior to use in the assessment. Although PISA is intended to measure authentic tasks, the PISA framework does not emphasize the use of existing, intact text. PISA is constrained in some ways by its international nature, as passages must be applicable across a wide range of cultures and languages. Therefore, while passages are selected to represent a range of texts and applicability in real-world settings, more manipulation and editing of passages is acceptable in PISA than in NAEP.
- PISA and NAEP measure similar cognitive skills, according to the cognitive dimension of the frameworks. Both measure students' ability to locate specifically stated information in a text, to make inferences and interpretations within and across text, and to evaluate or reflect on what they have read. PISA places slightly more

[1] Based on a review of approximately 70 percent of the passages on each assessment.

emphasis on the "locate" category and slightly less on the "reflect/evaluate" category than does NAEP at grade 8 and 12. Moreover, while the labels of the three categories used to define the cognitive dimension are similar, the panel's examination of the category descriptions and items reveal some differences in what is being measured.

- The panel examined PISA and NAEP items to determine if each would be comparably classified on the other assessment, according to the frameworks. For example, would a particular PISA item classified as integrate and interpret be similarly classified on NAEP (i.e., in the NAEP integrate and interpret category)? The panel found that about 90 percent of both NAEP eighth- and twelfth-grade items fit PISA's cognitive categories tightly and well (that is, could be comparably classified on PISA), whereas about 80 percent of PISA items fit the NAEP cognitive categories tightly and well; about 5 percent of items in each assessment were thought to not be appropriate for the other assessment in terms of what was being assessed. Although the panel members thought that most items could "fit" on the other assessment in terms of the framework category definitions, they also found that many items in each assessment were presented or formatted in ways that were not typical of or appropriate for the other assessment. Finally, while NAEP assesses "meaning vocabulary," that is, the meaning of words as they are used in the context of the particular passage, PISA does not include any items of this type.

Information about how the PISA mathematics and science assessments compare with the NAEP and Trends in International Mathematics and Science (TIMSS) mathematics and science assessments are available on the NCES website: http://nces.ed.gov/timss/pdf/Comparing_TIMSS_NAEP_%20PISA.pdf.

www.ed.gov

ies.ed.gov