ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

# User's Guide for the Program for International Student Assessment (PISA)

## 2006 Data Files and Database with United States Specific Variables

NCES 2009-055

U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

# User's Guide for the Program for International Student Assessment (PISA)

## 2006 Data Files and Database with United States Specific Variables

## April 2009

**Patricia J. Green**
**Deborah Herget**
**Jeffrey Rosen**
RTI International

**Holly Xie**
Project Officer
National Center for Education Statistics

NCES 2009-055

U.S. DEPARTMENT OF EDUCATION

**U.S. Department of Education**
Arne Duncan
*Secretary*

**Institute of Education Sciences**
Sue Betka
*Acting 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, and 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

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

**April 2009**

The NCES World Wide Web Home Page is http://nces.ed.gov.
The NCES World Wide Web Electronic Catalog is http://nces.ed.gov/pubsearch.

**Suggested Citation**
Green, P.J., Herget, D., and Rosen, J. (2009). *User's Guide for the Program for International Student Assessment (PISA): 2006 Data Files and Database with United States Specific Variables* (NCES 2009-055). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

**For ordering information on this report, write to**
U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794-1398
or call toll free 1-877-4ED-Pubs or order online at http://www.edpubs.org

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

# Contents

# List of Tables



## Appendix Tables

**Table** **Page**

**Table** **Page**

# List of Figures

**Figure** **Page**

# Chapter 1. A User's Guide to U.S. Data from PISA 2006

This user's guide is designed to provide researchers with an overview of the design and implementation of the 2006 Program for International Student Assessment (PISA), as well as with information on how to access the PISA 2006 data. This information is meant to supplement that presented in Organization for Economic Cooperation and Development (OECD) publications by describing those aspects of PISA 2006 that are unique to the United States.

## 1.1 Overview of PISA 2006

PISA is an assessment of 15-year-old students'[1] capabilities in reading literacy, mathematics literacy, and science literacy. PISA 2006 is the third administration of PISA. PISA was first administered in 2000 and is administered every 3 years. It is developed and administered under the auspices of the OECD, an intergovernmental organization of industrialized countries. In the United States, PISA is funded by the National Center for Education Statistics (NCES), a center within the Institute of Education Sciences of the U.S. Department of Education. In 2006, PISA was administered in 57 jurisdictions around the world, including all 30 OECD member countries.

PISA assesses all three subject areas each administration, but focuses on one of the three subject areas in order to provide in depth assessment information on the focal subject, as well as information about students' attitudes and experiences concerning the focal subject. In this third administration, PISA 2006, science literacy was the subject area assessed in depth. Reading literacy was the focal subject in 2000 and mathematics literacy was the focal subject in 2003.

PISA differs from other national and international assessments in some important ways. PISA does not focus explicitly on curricular outcomes but on the application of knowledge and skills in a real-life context. *Science literacy* refers to students' scientific knowledge and the use of that knowledge to identify questions, acquire new knowledge, explain scientific phenomena, and draw evidence-based conclusions about science-related issues; their understanding of the characteristic features of science as a form of human knowledge and inquiry; their awareness of how science and technology shape our material, intellectual, and cultural environments; and their willingness to engage in science-related issues—and with the ideas of science—as reflective citizens. *Reading literacy* refers to students' capacity to understand, use, and reflect on written texts to achieve their goals, develop their knowledge and potential, and participate in society. *Mathematical literacy* refers to students' capacity to identify and understand the role that mathematics plays in the world, make well-founded judgments, and use and engage with mathematics in ways that meet their needs as constructive, concerned, and reflective citizens.[2] A comparison of the PISA and NAEP assessments can be found in appendix A. In addition to assessing students' knowledge in skills in the three subject areas, PISA also assesses students' attitudes toward the focal subject of each administration (in 2006, this was attitudes toward science) and collects additional contextual information from students and school administrators. The PISA 2006 student questionnaire asked students about their family backgrounds, attitudes toward learning, attitudes toward science, and learning experiences in

[1] Age 15 was chosen because, in most countries, students are approaching the end of compulsory schooling at this point. Surveying students at this age assesses their academic ability after roughly 10 years of formal education.

[2] For more specifics on the science, mathematics, and reading domains, see *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD 2006).

and out of school. The school questionnaire asked administrators about schools' human and material resources, admission policies, decision-making processes, and staffing practices, among other aspects of schools' learning contexts.

## 1.2 PISA 2006 Assessment Design

Accurate and reliable measurement of the entire PISA assessment framework requires more assessment items than an individual student can take without unduly burdening the student. Therefore the assessment was developed using Item Response Theory methods that produce several item booklets, which are subsets of the overall pool of items. Each individual student takes a single booklet. The PISA 2006 assessment used a Balanced Incomplete Block (BIB) design in which there were 13 item booklets. Students' responses on individual booklets are then aggregated across booklets to produce overall scores for each country and subgroups of students within countries. No individual student takes enough items for the estimation of a reliable and accurate score—scores are reported only for countries or other groups of students (for example, U.S. females or males). For purposes of analyses, PISA datasets include five "plausible values" for each student for each overall subject area score and subscale score. These are to be used for analyses of the performance of groups of students and not for analyses at the individual student level. Moreover, PISA used a complex sampling design, in which a sample of schools was randomly selected from the population of U.S. schools with 15-year-old students and then students were randomly selected from the selected schools. Because students were selected using this complex design, analysts must use software and macros that have been specially developed for this type of sampling design. The software and macros are described in chapter 11.

## 1.3 How PISA 2006 Was Conducted in the United States

Countries were required to conduct PISA using a standardized set of procedures. The oversight body for PISA 2006 communicated these standards through various international meetings and a national project manager's manual (PISA Project Consortium 2005a). Each country was responsible for its own data collection, following the guidelines set forth by the PISA consortium. The PISA consortium is responsible for the design and implementation of the assessment and was led in PISA 2006 by the Australian Council for Educational Research (ACER). Other partners in the consortium include the National Institute for Educational Measurement in the Netherlands (Cito), the National Institute for Educational Policy Research in Japan (NIER), and Educational Testing Service (ETS) and Westat in the United States. Quality monitors were hired by the PISA consortium to observe the data collection to ensure that guidelines were followed.

The conduct of PISA 2006 in the United States involved sampling schools and students, recruiting schools and students to participate in the study, using the prescribed forms to document plans and outcomes, developing and distributing instruments, collecting the PISA 2006 data, scoring constructed response items, and processing the data. These activities were conducted by RTI International under contract to NCES. Final cleaning, scaling, and weighting of the data were completed for all countries by the PISA consortium.

# Chapter 2. Sampling Schools and Students

## 2.1 PISA 2006 Sample Design

Each country was required to follow international requirements for designing and selecting the sample as given in the PISA sampling manual (PISA Project Consortium 2005b). The sampling design defines the target population and sets the requirement for participation yields. The target population for PISA 2006 was 15-year-old students who attended educational institutions in grade 7 or higher (PISA Project Consortium 2005a). To obtain reliable estimates, each country was required to assess a minimum of 4,500 students from a minimum of 150 schools, using a self-weighting sample design. The school response rate target was 85 percent for all jurisdictions and the student response rate target was 80 percent. A minimum of 65 percent of schools from the original sample of schools were required to participate for a jurisdiction's data to be included in the international database. Jurisdictions were allowed to use replacement schools (selected during the sampling process) to increase the response rate once the 65 percent benchmark had been reached. The international guidelines define the response rate as the number of participating school (both original and substitute schools), over the total number of eligible original sample schools.

The design of the U.S. school sample for PISA 2006 was developed to achieve each of the international requirements set forth in the PISA sampling manual. The U.S. school sample is a stratified sample that consists of two stages and uses probability proportional to size (PPS). The measure of size used in the first stage was the expected number of eligible 15-year-old students in the school. At the second stage, a sample of 42 students was selected from each school regardless of size, or all eligible students if there were fewer than 42.

A list of schools for the U.S. sample was prepared using data from the 2003–04 Common Core of Data (CCD) and preliminary data from the 2003–04 Private School Universe Survey (PSS), two NCES surveys. These schools were stratified into two explicit groups: schools with large enrollments of 15-year-old students and schools with small enrollments of 15-year old students. The frame was implicitly stratified (i.e., sorted for sampling) by five categorical stratification variables: grade span of the school, control of school (public or private), region of the country, type of location relative to populous areas, and proportion of non-White students (above or below 15 percent). The last variable used for sorting within the implicit stratification was the estimated enrollment of 15-year-olds based on grade.

Schools were selected in the first stage with PPS, and students were sampled in the second stage, yielding overall equal probabilities of selection. This was the model used in PISA 2003, though not in PISA 2000. For PISA 2000, the U.S. school sample had a three-stage design, the first stage of which was the selection of a sample of geographic primary sampling units (PSUs). However, the sample was not clustered at the geographic level either in PISA 2006 or in PISA 2003. The change to a two-stage model was made in PISA 2003 to reduce the design effects observed in the 2000 data and to minimize respondent burden on individual districts by spreading it across school districts as much as possible.

Once the sample was drawn, it was loaded into KeyQuest, a software program written specifically for countries participating in PISA. KeyQuest was used to manage the sample, draw the

student sample, track participation, and produce verification reports used to clean the data in preparation for submitting the data file.

## 2.2 Selecting Substitute Schools

The international requirements for PISA state that each country must make every effort to obtain cooperation from the sampled schools, but they also recognize that it is not always possible to obtain cooperation. Thus, it is allowable to use substitute schools (also called replacement schools) as a means to avoid loss of sample size associated with school nonresponse. The international guidelines state that at least 65 percent of participating schools must be from the original sample; additional schools may be selected from the substitute schools identified during the sampling process after that minimum is reached. Each sampled school was assigned two substitute schools in the sampling frame. If the original sample school refused to participate, a substitute school was asked to participate.

Following the PISA guidelines, substitutes for noncooperating sampled schools were identified by assigning as substitute schools the schools that immediately preceded and followed the sampled school on the frame. The sampling frame was sorted by the stratification variables and by measure of size to ensure that any sampled school's substitutes had similar characteristics. There were several constraints on the assignment of substitutes. One sampled school was not allowed to replace another, and a given school could not be assigned to replace more than one sampled school. Furthermore, substitutes were required to be in the same implicit 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 was designated a first substitute and the other a second substitute. If an original school refused to participate, the first substitute was then contacted. If that school also refused to participate, the second school was then contacted.

## 2.3 The U.S. PISA 2006 Sample

The U.S. school sample for PISA 2006 consisted of 236 schools (from 44 states)[3] containing at least one 7th- through 12th-grade class. This number was increased from the international minimum requirement of 150 to offset school nonresponse and reduce design effects. Using data from the 2003–04 CCD and the 2003–04 PSS, the schools were selected with probability proportional to their estimated enrollment of 15-year-olds. There were 27 sampled schools identified as ineligible or closed, reducing the sample to 209 schools.

Participating schools provided lists of 15-year-old students, and a sample of 42 students was selected within each school in an equal probability sample. If a school had fewer than 42 eligible students enrolled, all eligible students were selected for PISA. 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 selection.

## 2.4 Exclusions to the Sample

The PISA consortium established rules for excluding students who were unable to take the test. These exclusions could occur at two points: at the school selection stage or at the student

[3] The "exclusion" of states from the sample is a result of the random selection process. Since the sample was drawn with probability proportional to size, states with fewer and smaller schools had a lower probability of being represented in the sample. If a state had a total measure of size equal to or less than the sampling interval, there was a non-zero probability that no schools would be selected from that state and that it would not be represented in the sample.

selection stage. In the United States, students were excluded at both stages. At the first stage, all students in schools classified as serving only students with disabilities were excluded. At the second stage of sampling students, students were excluded if they fell into any one of the three types of exclusion categories defined in the international guidelines:

- *Students with functional disabilities.* These were students with a moderate to severe permanent physical disability such that they could not perform in the PISA testing environment.
- *Students with intellectual disabilities.* These were students with a mental or emotional disability who had been tested as cognitively delayed or who were considered in the professional opinion of qualified staff to be cognitively delayed such that they could not perform in the PISA testing situation.
- *Students with insufficient language experience.* These were students who met the three criteria of (1) not being a native speaker in the assessment language, (2) having limited proficiency in the assessment language, and (3) having received less than a year of instruction in the assessment language. In the United States, English was the exclusive language of the assessment

Tables 1 and 2 present information on the total number of students sampled and the number excluded. In table 1, column 1 shows the total number of 15-year-olds according to the most recent information (2004) available at the time of sampling. Column 2 shows the number of students enrolled in schools that were excluded from the national desired target population from the sampling frame. Column 3 shows the size of the national target population after subtracting the students enrolled in excluded schools. Column 4 shows the percentage of students enrolled in excluded schools. It is obtained by dividing column 3 by column 2 and multiplying by 100.

In table 2, column 1 shows the number of students participating in PISA 2006 in the United States. Column 2 shows the weighted number of participating students, *i.e.*, the number of students in the national defined target population that the PISA sample represents (this number is calculated by summing the student weights for participants). Column 3 indicates the total number of students who were excluded at the student sampling stage, *i.e.*, the number of sampled students who were identified by school coordinators as falling into one of the 3 exclusion categories defined above. Column 4 indicates the weighted number of excluded students, *i.e.*, the overall number of students in the nationally defined target population represented by the number of students excluded from the sample (this number is calculated by summing the student weights for participants). Column 5 shows the percentage of students excluded at the student sampling stage. Column 6 shows the overall exclusion rate, which is the sum of the percentage of students enrolled in excluded schools and the percentage of sampled students excluded at the student sampling stage.

**Table 1. Total enrolled students, population, and school-level exclusions for the United States: PISA 2006**

| Total enrolled population of 15-year-olds at grade 7 or above (national target population) | Total students excluded at school sampling stage | Total in national desired target population after school-level exclusions and before within-school exclusions | School-level exclusion rate (percent) |
|---|---|---|---|
| 4,192,939 | 19,710 | 4,173,229 | 0.47 |

SOURCE: Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006: Science Competencies for Tomorrow's World. Volume 1: Analysis.* Paris: Author (annex A, table A2.1, page 349).

**Table 2. Participating students, student-level exclusions, and overall exclusions for the United States: PISA 2006**

| Number of participating students | Weighted number of participating students | Excluded students at student sampling stage | | Percent | |
|---|---|---|---|---|---|
| | | Number | Weighted number | Within-school exclusion rate | Overall exclusion rate |
| 5,611 | 3,578,040 | 254 | 142,517 | 3.83 | 4.28 |

SOURCE: Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006: Science Competencies for Tomorrow's World. Volume 1: Analysis*. Paris: Author (annex A, table A2.1, page 350).

International guidelines state that less than 5 percent of a country's target population should be excluded from the sample. As shown in tables 1 and 2, the U.S. exclusion rate of 4.28 percent fell within international guidelines, with approximately 0.47 percent of students excluded at the school sampling stage and 3.83 percent of students excluded at the student sampling stage.

# Chapter 3. Response Rates

## 3.1 Response Rate Targets

Each country participating in PISA 2006 was required to achieve minimum response rate standards in order to be included in international analyses and reports. Each country was required to collect data from at least 4,500 students from at least 150 schools. The minimum school response rate was 65 percent of sampled schools. Countries achieving at least a 65 percent school response rate could increase the school response rate through the use of substitute schools that were selected during sample selection.

Several rules were established for calculating response rates (PISA Project Consortium 2004). Schools with at least 50 percent of selected students participating were considered participants. A minimum student participation rate of 80 percent was required of each participating country. A student was considered a participant if he or she completed the test session, answering at least one item.

## 3.2 Response Rate Outcomes

The original U.S. PISA 2006 sample consisted of 209 eligible schools. A total of 166 schools participated, for a weighted and unweighted school response rate of 79 percent.[4] Of the 166 participating schools, 145 were original schools and 21 were substitute schools. The weighted school response rate before substitution was 69 percent. Each of the participating schools achieved over 50 percent student participation and was included in the response rate calculations.

A total of 6,796 U.S. students were selected to participate in PISA 2006. Of these students, 363 were ruled ineligible because they left the school or because their grade or birth date was out of the PISA 2006 range. An additional 254 students were excluded using the criteria described in section 2.4. Thus, a total of 6,179 eligible students remained in the sample. Of the 6,179 eligible students, 5,623 participated. During data processing 12 cases were deleted as ineligible because of out-of-range birthdates found after cleaning, leaving 5,611 cases in the final U.S. data file, for a weighted and unweighted student participation rate of 91 percent.

One administrator at each school was asked to complete the school questionnaire. Of the 166 schools whose students participated in PISA 2006, 163 completed the school questionnaire, for an unweighted school questionnaire response rate of 98 percent.

Detailed unweighted participation rates are provided in table 3.

[4] The 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 substitute schools that participated in the denominator as well as in the numerator, and to add to the denominator any substitute schools that were "hard" refusals. This would result in an overall school response rate of 67.5 percent.
In addition, the school response rate shown here were calculated by the PISA consortium and is based on the number of responding schools, weighted by enrollment, divided by the number of schools sampled, weighted by enrollment. NCES would typically calculate the weighted school response rate by summing the basic weights for responding schools and dividing that by the sum of the basic weights of all sampled schools.

**Table 3. Detailed unweighted U.S. participation rates: PISA 2006**

| PISA outcomes | Number | Percent |
|---|---|---|
| School sample | | |
| Total | 236 | 100.0 |
| Closed or ineligible | 27 | 11.4 |
| Total eligible | 209 | 88.6 |
| Overall school participation | | |
| Total eligible | 209 | 100.0 |
| Participating original schools | 145 | 69.4 |
| Participating substitute schools | 21 | 10.0 |
| Schools deleted during processing | 0 | 0.0 |
| Total participating schools | 166 | 79.4 |
| Session logistics | | |
| Total participating schools | 166 | 100.0 |
| Schools participating during school hours | 88 | 53.0 |
| Schools participating outside of school hours | 78 | 47.0 |
| Schools requiring follow-up sessions | 74 | 44.6 |
| Overall school questionnaire participation | | |
| Total participating schools | 166 | 100.0 |
| Completed school questionnaires—school questionnaire response rate | 163 | 98.2 |
| Student sample | | |
| Total | 6,796 | 100.0 |
| Ineligible due to age, grade, or enrollment status | 363 | 5.3 |
| Exclusions due to functional disability | 24 | .4 |
| Exclusions due to intellectual disability | 191 | 2.8 |
| Exclusions due to limited English proficiency | 39 | .6 |
| Total Exclusions | 254 | 3.7 |
| Total eligible | 6,179 | 90.9 |
| Overall student participation | | |
| Total eligible | 6,179 | 100.0 |
| Nonrespondents: refusals | 123 | 2.0 |
| Nonrespondents: absent | 433 | 7.0 |
| Completes during test day session | 5,030 | 81.4 |
| Completes during follow-up session | 593 | 9.6 |
| Total participants | 5,623 | 91.0 |
| Deleted during processing as ineligible | 12 | 0.03 |
| Total in the data file—final student participation rate | 5,611 | 90.8 |

SOURCE: U.S. Department of Education, National Center for Education Statistics, Program for International Student Assessment (PISA), 2006.

Final school response rates are presented in tables 4 and 5. Although the response rates met the minimum international requirement, they failed to meet the 85 percent minimum response rate required by NCES statistical standards. As a result, a bias analysis was conducted to determine if the characteristics of nonresponding schools differed from those of responding schools; the report can be found in appendix B.

**Table 4. U.S. school response rates before substitution: PISA 2006**

| Unweighted | | Weighted by enrollment | | Weighted participation rate before substitution (percent) |
|---|---|---|---|---|
| Number of responding schools | Number of responding and nonresponding schools | Number of students represented by responding schools | Number of students represented by sampled schools | |
| 145 | 209 | 2,689,741 | 3,901,131 | 68.95 |

SOURCE: Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006: Science Competencies for Tomorrow's World. Volume 1: Analysis.* Paris: Author (table A2.3, page 355).

**Table 5. U.S. school response rates after substitution: PISA 2006**

| Unweighted | | Weighted by enrollment | | Weighted participation rate after substitution (percent) |
|---|---|---|---|---|
| Number of responding schools | Number of responding and nonresponding schools | Number of students represented by responding schools | Number of students represented by sampled schools | |
| 166 | 209 | 3,085,548 | 3,901,521 | 79.09 |

SOURCE: Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006: Science Competencies for Tomorrow's World. Volume 1: Analysis.* Paris: Author (table A2.3, page 355).

Table 6 presents information on the final student response rates and weighted totals, using the student base weights. The U.S. response rate of 91 percent exceeded the international requirement of 80 percent. Response rates exceeded the international requirement of 50 percent in each school, so all participating schools and students are included in the response rates presented here.

**Table 6. Final U.S. sample—students within schools after substitution: PISA 2006**

| Unweighted | | Weighted | | |
|---|---|---|---|---|
| Number of students assessed | Number of students sampled (assessed, absent and refused) | Number of students assessed | Number of students sampled (assessed, absent and refused) | Student participation rate after substitution (percent) |
| 5,611 | 6,179 | 2,589,680 | 2,845,841 | 91.00 |

SOURCE: Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006: Science Competencies for Tomorrow's World. Volume 1: Analysis.* Paris: Author (Table A2.3, page 355).

# Chapter 4. Recruitment of Schools and Students

International standards set response rate targets for PISA 2006 to be at least 65 percent for original sample schools and 80 percent for students. Recruitment, therefore, became a critical activity to ensure the success of PISA in the United States.

## 4.1 Task Force and Focus Groups

After experiencing difficulties achieving high levels of school participation in recent international studies, NCES and the Education Statistics Services Institute (ESSI)[5] convened a task force prior to the start of PISA 2006 to make recommendations to improve participation rates in international studies. The plan for recruiting schools was based partially on the recommendations of the task force. For the PISA 2006 recruitment effort, the following recommendations were followed:

- Begin the recruitment of schools for PISA at least a year prior to the scheduled data collection.
- Approach selected schools for participation in PISA directly and send an informational letter to states and public school districts alerting them that schools in their state/district had been selected.
- Employ in-person contacts with the schools, emphasizing personal contact with schools that decline to participate in order to convert the refusals.

NCES also conducted two focus groups with school principals prior to the start of recruitment for the PISA 2006 main study. The sessions with focus groups were intended to learn more about what might facilitate participation in this voluntary, low-stakes assessment and what would help keep schools engaged over the summer months, since recruitment would take place in the school year prior to the fall data collection. The principals were also asked about the feasibility of conducting PISA 2006 outside school hours, either after school or on a Saturday, and the incentive amount that would be necessary to obtain a high student response.

School principals were not receptive to using school time to conduct PISA but were cautiously optimistic about conducting sessions outside of school hours. They felt that a $50 incentive would be required for students to give up their afterschool or Saturday scheduled activities. Principals also felt that conducting a conference explaining the global context of science literacy would help keep the schools interested through the summer.

## 4.2 Contacting States, Districts, and Schools

Recruitment began a year before the start of data collection. It was hoped that the lead time would allow schools to plan for the resources required for participation within the study's schedule constraints, allow schools ample time to initiate and complete any internal review procedures they felt necessary, and allow the contractor time to work with the schools to resolve any potential roadblocks to their participation. By starting a year before data collection, it was hoped that a higher response rate from schools could be achieved.

[5] ESSI provides developmental, analytical, methodological, and operational support to NCES.

Prior to conducting a field test, endorsements were secured from 14 organizations for PISA 2006 and the 2006 Progress in International Reading Literacy Study (PIRLS 2006), which were conducted concurrently. Copies of the endorsement letters were included in recruiting materials. See appendix C for a list of the 14 organizations that endorsed PISA 2006.

### 4.2.1 State and District Notification

Most recruiting efforts approach officials in states, districts, and schools (in that order) to obtain participation. The NCES/Education Statistics Services Institute (ESSI) task force recommended that less resistance might be encountered if states and districts were notified in writing and if formal recruitment was begun directly with the schools.

Informational materials were mailed to each of the 44 chief state school officers (CSSOs) from states selected for the PISA 2006 main study. The package included a lead letter from NCES, a study fact sheet, a sample copy of the implicit parental consent form, and copies of the endorsement letters. The letter to the CSSO included a toll-free number in case the state official wanted to call with questions. Copies of the packet were also sent to the state assessment coordinator for each state. See appendix D for copies of the recruitment materials.

On the same date, informational materials were mailed to each of the districts or dioceses that had schools in the original sample. The package contained the same materials included in the state mailing, though the cover letter to the districts included the names of the schools sampled within their district. District letters were sent on a flow basis to the districts contributing one or more substitute schools to the sample. The district superintendents received a package that was identical to the one received by superintendents of districts with original schools. A total of 210 districts were sent informational mailings.

There were a handful of districts for which it was known, from prior experience, that a formal application and district approval procedure would need to be followed. In those situations, the necessary documentation was included in the initial mailing to the district. Several other districts contacted the data collection contractor to explain their application or approval procedures. In total, 23 districts required the preparation and submission of formal applications to conduct research in the district's schools, and 5 districts had other processes that needed to be followed before the school(s) in those districts could be contacted.

### 4.2.2 School Contacting

Schools were contacted after the districts were notified and, when applicable, after district applications to conduct research were approved. The school mailing was addressed to the principal and contained a cover letter, a study fact sheet, a sample copy of our implicit parental consent form, endorsement letters, and a study brochure. Several days after the package was sent, the school was contacted by telephone. A copy of the school recruitment letter is provided in appendix D.

The principal at each participating school designated one person to serve as the school coordinator for PISA. School coordinators were asked to work with project staff to coordinate the logistics of the test session and to ensure high student response. An honorarium was provided to the school coordinator, with the exact amount dependent upon student attendance at the test session.

## 4.3 Efforts to Gain School Cooperation in the United States

To achieve the minimum required school response rate of 65 percent, the United States introduced several innovations. First, the United States (and two other countries) moved the testing period from spring 2006 to fall 2006; second, the United States sponsored a conference for schools participating in PISA 2006; and third, the United States arranged to administer the test outside of normal school hours in some schools that wished to participate but did not want to do so during instructional time. These efforts are described below.

The United States was one of three countries to select fall 2006, rather than spring 2006, for its data collection period. The United States made this choice to avoid conflicting with mandatory high-stakes testing that often occurs in the spring, based upon the PISA 2003 experience. In 2003, the United States was unable to achieve the 65 percent school response rate in the spring and was allowed to extend data collection into the fall. Careful analyses of the 2003 data indicated no differences between average scores of students who completed the assessment in the fall and in the spring. In 2006, the international consortium refused to allow countries to administer the test over such an extended period, but did allow countries to choose a fall administration if the participating country believed that school response rates would improve. Three countries (the United States, the United Kingdom, and Bulgaria) moved the test date to the fall; consequently, the range of eligible birthdates in those countries was adjusted to ensure that the mean age remained consistent across all countries. In the United States, students born between July 1, 1990, and June 30, 1991, were eligible to participate.

A conference entitled "The Global Context of Scientific Literacy" was convened in Washington, DC, from July 31 through August 1, 2006. The purpose of the conference was to keep the schools engaged in the study during the summer months and avert the possibility that they would renege on their commitment to participate in the fall. One representative from each participating school was invited to attend. At the conference, national project managers from other participating countries (Australia, Canada, Hong Kong, New Zealand, and the United Kingdom) described why PISA is important in their country and how the data are used. Individuals who use international education data in the United States demonstrated how they use the data and why international data are important in the United States. PISA project staff also prepared a presentation on what school staff and students should expect when PISA is administered in their school. The conference was successful in building enthusiasm for the study, and only one of the schools that sent a representative to the conference dropped out before PISA was administered.

Schools were offered the option of conducting the assessment after school hours or on a Saturday. This option was offered only as a refusal conversion tool and not as part of the initial recruitment materials. Of the 166 participating schools, 88 schools conducted the session during school hours, 4 conducted the session after school, and 74 participated on a Saturday. The student response rate was 91 percent during school hours and 90 percent in schools where PISA was administered after school or on a Saturday. Analyses were conducted comparing the performance of students who took the test during the regular school day with those who took the exam after school or on a Saturday. No measurable differences were found between the two groups.

## 4.4 Recruiting Parents and Students in the United States

Once the student sample was selected within a school, PISA staff worked with the school coordinator to obtain parental consent. Schools were given a choice between two types of parental permission forms in PISA 2006: implicit or explicit. The majority of schools, 74 percent, chose to

use implicit consent, in which parents returned the form only when they objected to their 15-year-old's participation. The remaining 26 percent required explicit or written parental consent. Parental consent forms were provided in additional languages upon request by the school.

## 4.5 Student-Level Incentives

Incentives were offered at the student level in an effort to offset some of the stress associated with test taking and to motivate the school to participate by giving something back to the students. Student-level incentives also served as a token of appreciation for participating in the study. Students received $15 for participating during regular school hours. If the school required the administration to take place outside of school hours (after school or on a Saturday), students received $50. In addition to the monetary incentive, each participating student received a certificate to commemorate that he or she had represented the United States in PISA 2006.

# Chapter 5. Sampling and Data Collection Forms

The forms discussed in this section were instrumental in the 2006 U.S. data collection. With two exceptions (see sections 5.3 and 5.5, below), the forms are standard international forms developed by PISA for use in all countries (PISA Project Consortium 2005c).

## 5.1 Student Listing Form

To draw the student sample from each school, a Student Listing Form was provided to schools to list all of the age-eligible students enrolled at the school. The completed form was used to draw the student sample. Explicit instructions and guidelines for completing the form were provided, and electronic submission was encouraged. Although the basis of the Student Listing Form is an international form, minor adaptations were made to the instructions to cover needs specific to the U.S. data collection.

## 5.2 Student Tracking Form

The Student Tracking Form was used to track the participation and nonparticipation status of each sampled student at the school. This form remained unchanged from the international version and was output from the KeyQuest software as required. All identification and sampling information (the upper portion of the form) was filled in by the KeyQuest program, as well as the identification number, line number, student name, and study program. These data were obtained through communication with the school coordinator and from the list of eligible students provided by all participating schools. The remaining columns were completed by the test administrator to record exclusions and students with special education needs, as well as the participation status of each of the other students.

### 5.2.1 Instructions for Defining Students With Special Education Needs

The Student Tracking Form contains a column for indicating students with special education needs (SEN). Instructions to assist in defining these students—i.e., those with intellectual or functional disabilities and those with limited English proficiency (LEP)—were provided to each school coordinator; test administrators were available to assist in this process if needed. Students with SEN needed to be coded in the Student Tracking Form but were not to be automatically excluded from the assessment.

### 5.2.2 Instructions for Including/Excluding Students

Exclusion guidelines state that overall estimated student exclusions should be under 5 percent (OECD 2009, p. 66). The Student Tracking Form records information on each student's inclusion/exclusion status. A student identified with SEN was not automatically excluded from the assessment. An exclusion code was applied only if the SEN existed to the degree that a school official or the test administrator deemed the student unable to perform in the PISA testing situation. The instructions for including/excluding students provided clear directions for applying the correct codes for each student listed on the Student Tracking Form.

### 5.2.3 Student Exclusions in PISA 2006

Of the 6,796 students identified in the U.S. PISA 2006 sample, schools excluded 254 from the assessment using the international exclusion criteria supplied to them. The breakdown of excluded students was as follows: functional disability, 24; intellectual disability, 191; and LEP, 39. The resulting weighted exclusion rate, including exclusions at the school-sampling stage, was 4.3 percent. This exclusion rate was lower than the 7.5 percent exclusion rate reported for PISA in 2003, in line with the 4 percent rate reported for PISA in 2000, and also in line with the international exclusion guidelines discussed in section 2.4.

## 5.3 School Logistics Form

The School Logistics Form was developed for use in the United States to provide the test administrators with detailed information about the participating school. Information about prior contacts with the school, procedures for obtaining parental consent, and scheduling issues were included on this form.

## 5.4 Session Report Form

The Session Report Form is an international form that was used to capture information about each assessment session. The test administrator completed most of this form during the session by recording session timing, student behavior, any disruptions that may have occurred during the session, and any information on specific assessment booklet or questionnaire items that may have been problematic.

## 5.5 Return Shipment Form

The Return Shipment Form is a U.S. form that was used for receipt control. It provided a record of the materials returned to RTI and guidelines on packing and shipping procedures for the test administrator to follow.

# Chapter 6. Instrument Development and Distribution

## 6.1 Test Development

The development of the PISA 2006 assessment instruments was an interactive process among the PISA consortium, various expert committees, and Organization for Economic Cooperation and Development (OECD) members. The assessment was developed by international experts and PISA consortium test developers; it included items submitted by participating countries and items developed by the consortium's test developers. Representatives of each country reviewed the items for possible bias and for relevance to PISA's goals. The intention was to reflect in the assessment the national, cultural, and linguistic variety of the OECD countries. Science items were field tested in 2005 in each country to examine their psychometric properties and identify any problematic items. Following the field test, a number of statistics were reviewed for each item for each country; including percent correct, item difficulty, item discrimination, and gender differences. Items that worked differently across countries were deleted.

The final assessment consisted of 140 science items, 48 mathematics items, and 28 reading items, allocated to 13 test booklets. Each booklet was made up of four test clusters. There were seven science clusters (S1–S7), four mathematics clusters (M1–M4), and two reading clusters (R1–R2). The clusters were allocated in a rotated design to the 13 booklets. The average number of items per cluster was 20 for science, 12 for mathematics, and 14 for reading. Each cluster was designed to average 30 minutes of test material. In addition to the cognitive assessment, students also received a 30-minute questionnaire designed to provide information about their backgrounds, their attitudes, and their experiences in school. More detailed information on test development can be found in chapter 2 of the international technical report (OECD 2009).

The booklet rotation was assigned to students using software developed by the PISA consortium. Labels were generated and affixed to each test booklet and questionnaire to provide the student identification and booklet number. Booklets were then bundled by school and shipped to the test administrator at the school a week prior to the session.

## 6.2 Test Printing

The data collection contractor, RTI International, made an error printing the final test booklets in the United States, and the pagination of the booklets was consistently off by one page. The international consortium intended for the first page to be printed on the inside of the cover; in the United States it was typically printed on the first page of plain white paper. As a result, some of the instructions in the reading section were incorrect. In some passages, students were incorrectly instructed to refer to the passage on the "opposite page" when the passage now appeared on the previous page. Because of the small number of items in the reading section, it was not possible to recalibrate the score to exclude the affected items. As a consequence, ACER determined that the U.S. reading literacy data were invalid. No incorrect page references appeared in the mathematics or science sections of the booklets. However, in some instances, mathematics and science items may have been more difficult because the questions required information provided previously that now required the students to turn back a page. In a few instances, items may have been somewhat easier because of the pagination. ACER examined the potential impact of this on the mathematics and

science scales and estimated the impact on scores would fall within the equating error of the scale. Therefore, the original scales were retained using the results from all mathematics and science items.

# Chapter 7. Overview of Field Operations

## 7.1 Recruitment of Field Supervisors and Test Administrators

Four field supervisors (FSs) and 35 test administrators (TAs) were hired to work on the PISA 2006 main study in the United States. Each TA was assigned to one of the four FSs, who coordinated and monitored the TA's work.

## 7.2 Training of Field Supervisors and Test Administrators

Prior to training, each TA was mailed a copy of the test administrator manual (PISA Project Consortium 2005c) and a home study exercise. The test administrator manual was adapted from the international version to include procedures specific to conducting the study in the United States.

The TAs were instructed to read the manual prior to training and complete the home study exercise to be turned in on the first day of training. Project staff conducted training in Raleigh, North Carolina, on August 16–18, 2006. Each FS and TA signed a confidentiality agreement and an affidavit of nondisclosure at the beginning of training.

At the end of the training, project staff certified the TAs to ensure that they were prepared to administer the PISA 2006 test in the schools. The TAs were certified on their ability to answer common questions about PISA 2006, read the PISA Introduction Script verbatim, understand how to read the School Logistics Form, complete a Student Tracking Form exercise, and complete a Session Report Form.

## 7.3 Data Collection Procedures

The international school coordinator manual (PISA Project Consortium 2005d) was adapted to include procedures for the United States. School coordinators (SCs) were asked to provide a list of all enrolled students in grade 7 or higher with birth dates between July 1, 1990, and June 30, 1991. The student lists were used to sample 42 students per school in an effort to gain participation from an average of 35 eligible students per school.

SCs received Student Tracking Forms and the school questionnaire prior to the scheduled session. Parental consent forms and informational materials were mailed directly to parents, though SCs received copies to aid in obtaining student participation. The test booklets, questionnaires, and other administration materials (e.g., pencils, calculators) were sent directly to the TAs prior to the scheduled session. Each test booklet and questionnaire was labeled before it was sent to the TAs, who brought the materials with them to the schools. The TAs arrived at each school about one hour prior to the scheduled administration to set up for the session.

The main data collection period ran from September 25 through November 22, 2006. On the test day at each school, the TA checked in with the SC and collected any parental consent forms that had come in. In schools requiring explicit parental consent, the TA checked the Student Tracking Form to make sure that only students who had returned signed permission forms were allowed to participate. For all schools, the TA made sure that no one for whom the school received a parental refusal was allowed to participate unless the parent had rescinded that decision in writing.

The student response rate was 91.5 percent when implicit consent was used and 88.3 percent when explicit consent was required.

The TA distributed the labeled test booklets, questionnaires, and pencils to each student. Prior to the start of the test, the TA read to the students the PISA Introduction Script describing the study, giving the elements of informed consent, and giving instructions for completing the questionnaire and test.

The students were randomly assigned one of 13 test booklets. There was one version of the student questionnaire. The total amount of time needed to complete the material was about 3 hours, including time for distributing materials and reading instructions, the 2-hour test with a short break after the first hour, and the 30-minute student questionnaire. After completing the session, each participating student received a certificate stating that he or she had represented the United States in PISA 2006.

At the conclusion of the testing, the TA determined whether a makeup session was necessary. A makeup session was held in each school that did not achieve at least 85 percent student participation. Makeup sessions were conducted at 73 of the 166 schools.

## 7.4 School Questionnaire

In addition to the student component, PISA 2006 included a school questionnaire, which one school administrator was asked to complete. Although it was requested that the principal complete the school questionnaire, any knowledgeable school staff member could complete most of it. The principals, or their designees, were asked to complete the questionnaire at their convenience between the time of its receipt and the test day. The TAs collected the completed questionnaire on the test day, but if it had been lost or had not been completed, they provided a second questionnaire and asked the principal or designee to complete it while the test was in session. Most of the questionnaires were completed either before or during the student test administration. For school administrators who did not complete their questionnaire before the end of the test day session, the TAs left a preprinted Federal Express label for its return, and project staff prompted the administrators to complete and return it. Ninety-eight percent of school questionnaires were returned.

## 7.5 Quality Assurance Procedures with Field Staff

FSs maintained frequent communication with their TAs and closely monitored the quality of their work. Once a TA returned from a test session, he or she contacted the FS to report on the outcomes of the session. Project staff monitored the outcomes of the session and compared the information reported by the TA against the completed test booklets and questionnaires returned. Project staff and field staff used the Field Reporting System (FRS) to record and monitor the participation status of each student.

When the completed test booklets and questionnaires arrived, data receipt clerks removed the completed materials from their packaging and reviewed them closely against the information recorded on the Student Tracking Form. The materials were then separated by booklet type and logged into the Data Receipt System before being batched and stored in a secure location. Information recorded in the two systems (Data Receipt System and FRS) was compared to ensure that the materials expected to be received (based on participation codes recorded in the FRS) were actually logged in the Data Receipt System. The FRS generated reports to identify any discrepancies between what the FS recorded in the FRS and what was sent in by the TA. The primary types of

discrepancies the reports would indicate were missing booklets, a completed booklet received for a student who was recorded as a nonparticipant, or a refusal form received for a participant. The FS followed up with the TA in a timely fashion to resolve any discrepancies, and all discrepancies were resolved by the end of the data collection period.

After schools completed the test administration, project staff conducted verification reinterviews with school coordinators and principals to verify that the administration had gone well. Information about it was entered into the Field Reporting System, and feedback was provided to the TAs.

## 7.6 Data Entry

All participating countries were required to perform data entry on each of the test booklets and then re-enter 100 of each booklet type to assess keyer reliability. In the United States, four data entry staff were trained to key in the test booklets, questionnaires, and reliability sheets (the sheets that recorded the codes assigned by second, third, and fourth scorers; see chapter 8 for more information on coding reliability procedures); data re-entry on the test booklets was performed by a data entry supervisor.

KeyQuest included a data entry component into which the test booklets, reliability sheets and questionnaires were to be keyed. During the field test, KeyQuest became extremely slow and would frequently lock up when multiple users were entering data simultaneously. As a result, a new user interface was designed for data entry that loaded data directly into KeyQuest.

## 7.7 Data Cleaning and Validation Reports

KeyQuest included a series of validation reports to be run on the data file to identify issues with the data. All issues had to be resolved or documented before the data were submitted to the PISA consortium. To facilitate a quick turnaround between the end of data collection and the deadline for data submission, these reports were run on a flow basis throughout the data entry process. Once all of the test booklets and questionnaires had been keyed, the validation reports were run for a final time to resolve any remaining issues. Problem resolution consisted of reviewing the Student Tracking Forms, test administration forms, test booklets, and questionnaires to resolve discrepancies that appeared in the validation reports. The U.S. data collection contractor cleaned as many discrepancies as possible and annotated the remaining issues, since, at times, items identified as problematic in the reports actually provided an accurate account of the test administration at the school.

The data file and annotated issues were sent to the PISA international coordinating center, ACER, in January 2007. ACER cleaned the data and sent inquiries as needed to resolve issues. Once data cleaning was completed, the data files from all participating countries were merged by ACER to produce an international data file.

# Chapter 8. Scoring and Coding

The scoring for PISA 2006 was conducted by CTB/McGraw-Hill (CTB), a subcontractor to RTI. CTB supervisors were responsible for hiring and training staff, ensuring quality control procedures were in place, and reporting outcomes back to RTI.

The PISA assessments included both multiple-choice and constructed-response items. The science literacy assessment included 41 constructed-response items, and the mathematics literacy assessment included 27 constructed-response items. All constructed-response items were scored by staff specifically trained in PISA scoring.

The process of scoring began with an international training session held in February 2006. The U.S. project director and two CTB supervisors attended this training session, where they received complete introductions to and explanations of all the PISA 2006 items. The items were thoroughly reviewed, with close attention paid to the specific student abilities being assessed. This training provided the supervisors with the knowledge they would need to hire and train the most appropriate scoring staff.

## 8.1 Staffing

Scoring staff were selected on the basis of their educational backgrounds and their prior experience with similar assessment projects. Each scorer signed the PISA 2006 confidentiality agreement.

## 8.2 Training

The next step in the process was the training of the scorers. The international consortium prepared a coding guide to be used by all countries, and staff from each country were trained in using the materials at the international training. The two supervisors from the United States who attended the international training session in February 2006 conducted training sessions for U.S. scoring staff. One supervisor trained the 16 scorers and two team leaders on the science and mathematics clusters; the other trained the 4 scorers and one team leader on the reading clusters. On the first day of training, the staff reviewed all test booklets and completed the test items on their own. After completing this task, the trainer began the review of each item's coding guidelines in the first cluster (training was conducted by cluster) and answered group questions. The training sessions included discussions of the example responses and papers in the coding guides provided by the PISA consortium.

One important activity during training was a review of item queries from the scorer query service established by the PISA consortium. Through this review, questions and comments regarding scoring and coding were collected and disseminated internationally. The U.S. training included regular reviews of the item queries. PISA 2006 included a series of linked items that also appeared in PISA 2003. Scorers completed a trend reliability exercise in which they scored actual 2003 student responses to items that appeared again in PISA 2006. To give scorers a sense of their own accuracy, their scores were compared with those of an expert scorer from PISA 2003. The trend reliability exercise also gave supervisors a sense of how reliably the 2006 scorers were scoring.

In sum, the training process consisted of a review of the coding guidelines for each cluster, practice with sample items, discussion of issues and questions, a review of item queries, and scoring of the trend reliability items. After the training process was completed for each cluster, the teams moved directly into the scoring of that cluster in the 2006 booklets.

## 8.3 Scoring Process and Quality Control

Procedures for scoring were provided in the *National Project Manager's Manual* (PISA Project Consortium 2005a, paragraphs 317 through 361). The manual provided guidelines for all aspects of the scoring process and suggested a schedule of 6.75 hours per day of scoring.

The primary quality control procedure during scoring consisted of read-behind reports on the completed work of each scorer. In a read-behind, the trainer, supervisor, or team leader rescored recently completed items. Team leaders completed read-behinds on the scorers, and supervisors and trainers completed read-behinds on both the scorers and team leaders. When scoring discrepancies occurred, they were discussed and resolved. Scorers who struggled during read-behind checks were coached by the team leader and supervisor and their work was continually reviewed. During scoring, the read-behind quality measures were calculated and reported daily. The complete read-behind reports can be found in appendix E, table E-1 (science) and table E-2 (mathematics).

## 8.4 Main Reliability Study

The design and process for the main scoring reliability study was laid out in the national project manager's manual (PISA Project Consortium 2005a). Although alternative designs were allowed, the United States adopted the standard multiple-coding design specified in the manual. Six hundred booklets were designated for multiple coding and were selected systematically from all student test booklets. Specifically, every fourth booklet from booklet types 1, 3, 5, 6, 8, and 10 was selected for multiple coding. Because test items were not identically laid out in each booklet type and clusters appeared at different places across the test booklets, selection from these 6 booklet types ensured that items from all mathematics, science and reading clusters were multiple coded. Multiple coding was blind. In other words, no multiple scorer was aware of scores given by a previous scorer. The booklets to be multiple coded were rotated in a manner that ensured coding by four independent scorers.

The outcome used to assess reliability was the percent agreement among the four scorers. When all scorers agreed on a score for a specific item, that instance of the item was coded as having complete agreement. When there was disagreement, such as when one scorer reported a different code for a specific item, that item was said to lack complete agreement and was coded as such. When it is reported, for example, that item M421Q01 has a final percent agreement of 92.73 percent, it can be interpreted as follows: in all instances where M421Q01 appeared in the booklets selected for multiple coding, all four scorers agreed on a score in 92.73 percent of the cases. Table E-3 in appendix E reports percent agreement for items selected for multiple coding. Agreement across these items is generally high, but what constitutes high (and low) agreement is essentially arbitrary.

The percent agreement approach is a rudimentary method of assessing inter-rater reliability and has shortcomings, one of which is an inability to control for chance agreement. However, the purpose of this task was to provide study managers with some ability to identify problem items during the scoring process. The multiple-coding design as implemented by the United States was not one in which reliability data were simply reported at the end of scoring. Instead, analysts reviewed

reliability data periodically throughout the scoring process. Early in the scoring process, a few items were found to have particularly low percent agreement (below 75 percent)[6] and were flagged for further examination. Scorers went through a new training on these low agreement items and then rescored all instances of the item in the booklets already scored, after which percent agreement was calculated again. In all cases, this resulted in higher percent agreements for the flagged items. The inter-rater reliability scores were checked again later and no major issues were found. Again, for study managers, this multiple-coding process was treated as quality control and as a process by which items with questionable reliability could be identified. A more sophisticated reliability analysis is reported in the international technical report (OECD 2009).

## 8.5 Trend Reliability Study

As stated previously, PISA 2006 contained a number of mathematics, science, and reading items that also appeared in PISA 2003. Primarily, trend items were included in PISA 2006 to assess changes in student learning over time. However, they can also be used for the critical task of assessing the *stability of scoring* of the constructed-response link items from 2003 to 2006. If the process of scoring or if the behavior of scorers changes over time, it becomes much more difficult to view changes over time in student scores. Thus, the trend reliability study conducted by the United States was designed to check for stable scoring. Table 7 lists the linked items in PISA 2006 in mathematics and science that were used for the trend scoring exercise.

**Table 7. Linked items in PISA 2006**

| Unit name | Item ID | Unit name | Item ID |
|---|---|---|---|
| Height | M421q01 | Good Vibrations | S131q02; S131q04 |
| Running Tracks | M406q01; M406q02; M406q03 | Algae | S268q02 |
| Carbon Dioxide | M828q01 | Earth's Temperature | S269q01; S269q03 |
| Population Pyramids | M155q02; M155q03 | Water | S304q01; S304q03a; S394q03b |
| Greenhouse | S114q03; S114q04; S114q05 | Milk | S326q01; S326q02 |

SOURCE: U.S. Department of Education, National Center for Education Statistics, Program for International Student Assessment (PISA), 2006.

### 8.5.1 Selecting Linked Items for Trend Scoring

A subset of 2003 student responses was used for the trend study. Thus, a large non-random sample (a random or systematic selection of items was not possible) was drawn from 2003 booklets 4, 8, and 10 to come up with a feasible number of items for 2006 scorers to rescore. Drawing items from these three booklets ensured that items from all clusters would be included in the analysis. All linked items from the selected booklets were used in the trend scoring study, which resulted in 125 science items and 68 mathematics items.

### 8.5.2 Trend Scoring Process

Trend scoring was conducted during scorer training. Each of the 16 mathematics and science scorers and 4 reading scorers were provided actual student responses to 2003 linked items.

[6] The percent agreement benchmark was set at 75 percent following reliability procedures put in place for PIRLS (see PIRLS 2001 Technical Report, Martin, M.O., Mullis, I.V.S., & Kennedy, A.M. (Eds.) (2003), Chestnut Hill, MA: Boston College). Again though, the placement of the benchmark is essentially arbitrary and the literature does not provide much guidance on acceptable levels of percent agreement, as it does for acceptable values of, for example, the inter-class correlation coefficient.

The booklets were organized by cluster to work seamlessly within the existing training processes. Immediately after training on a particular cluster, the scorers would rescore selected student responses to 2003 items in that cluster. In most cases, the scorers would rescore roughly 6 to 10 actual student responses to each item in each cluster. The scores would then be compared with scores given in 2003 by an expert scorer.

### 8.5.3 Outcomes

The percentage agreement between the 2003 expert scorer and the 2006 scorer was used as the outcome measure. Tables 8 and 9 report the final outcomes for the science and mathematics items (respectively) in the trend reliability study.

**Table 8. Percent agreement on U.S. science trend items, by scorer ID: PISA 2006**

| Scorer ID | Number reviewed | Number agreed | Percent agreed |
|---|---|---|---|
| Total | 2000 | 1809 | 90.5 |
| 501 | 125 | 106 | 84.8 |
| 502 | 125 | 112 | 89.6 |
| 503 | 125 | 114 | 91.2 |
| 504 | 125 | 114 | 91.2 |
| 505 | 125 | 112 | 89.6 |
| 506 | 125 | 111 | 88.8 |
| 507 | 125 | 115 | 92.0 |
| 508 | 125 | 114 | 91.2 |
| 509 | 125 | 118 | 94.4 |
| 510 | 125 | 118 | 94.4 |
| 511 | 125 | 111 | 88.8 |
| 512 | 125 | 117 | 93.6 |
| 513 | 125 | 113 | 90.4 |
| 514 | 125 | 113 | 90.4 |
| 515 | 125 | 109 | 87.2 |
| 516 | 125 | 112 | 89.6 |

NOTE: "Percent agreement" indicates the agreement between a scorer and an evaluator on a score given to a constructed response item. For example, if an evaluator checked 30 items by a scorer ("Reviewed" column) and agreed with the code given to open-ended responses for 26 of those items ("Agree" column), a percent agreement of 86.7 would result.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Program for International Student Assessment (PISA) 2006.

**Table 9. Percent agreement on U.S. mathematics trend items, by scorer ID: PISA 2006**

| Scorer ID | Number reviewed | Number agreed | Percent agreed |
|---|---|---|---|
| Total | 1,088 | 986 | 90.6 |
| | | | |
| 501 | 68 | 61 | 89.7 |
| 502 | 68 | 63 | 92.6 |
| 503 | 68 | 58 | 85.3 |
| 504 | 68 | 62 | 91.2 |
| 505 | 68 | 61 | 89.7 |
| 506 | 68 | 63 | 92.6 |
| 507 | 68 | 62 | 91.2 |
| 508 | 68 | 63 | 92.6 |
| 509 | 68 | 62 | 91.2 |
| 510 | 68 | 62 | 91.2 |
| 511 | 68 | 62 | 91.2 |
| 512 | 68 | 62 | 91.2 |
| 513 | 68 | 60 | 88.2 |
| 514 | 68 | 59 | 86.8 |
| 515 | 68 | 65 | 95.6 |
| 516 | 68 | 61 | 89.7 |

NOTE: "Percent agreement" indicates the agreement between a scorer and an evaluator on a score given to a constructed response item. For example, if an evaluator checked 30 items by a scorer ("Reviewed" column) and agreed with the code given to open-ended responses for 26 of those items ("Agree" column), a percent agreement of 86.7 would result.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Program for International Student Assessment (PISA) 2006.

# Chapter 9. Processing, Scaling and Weighting

## 9.1 International Data File Cleaning and Editing

After all scoring and data entry activities were completed, the data were copied back into KeyQuest. The validation reports were then run, and issues were resolved by going back to the original questionnaires and forms to examine inconsistencies. When all inconsistencies had been resolved or documented, the files were transferred to ACER.

ACER cleaned each of the national data files to ensure that data cleaning was standardized among all participating countries. 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 to anomalies. ACER then compiled background univariate statistics and preliminary classical and Rasch item analysis. Detailed information on the entire data cleaning process can be found in chapter 10 of the OECD *PISA 2006 Technical Report* (OECD 2009).

## 9.2 Missing Data

Unlike other NCES studies, PISA does not impute missing information for questionnaire variables. The international database and the U.S. database contain four kinds of missing data codes that are used across all countries. "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 coded "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 2006 database.

## 9.3 Scaling

The data from all participating countries was scaled by ACER using a mixed-coefficients multinomial logit model. Detailed information on scaling procedures can be found in chapters 9, 12, and 16 of the international technical report (OECD 2009).

## 9.4 Weights for U.S. Data

The use of sampling weights is necessary for the computation of statistically sound, nationally representative estimates. 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 jurisdictions participating in PISA 2006 was carried out by Westat, as part of the PISA consortium.

The internationally defined weighting specifications for PISA 2006 included two base weights and five adjustments. The school base weight was defined as the reciprocal of the school's probability of selection. (For substitute 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.

These base weights were then adjusted for school and student nonresponse. The school nonresponse adjustment was done individually for each jurisdiction using implicit and explicit strata defined as part of the sample design. In the case of the United States, three variables were used: school control, census region, and community type. The student nonresponse adjustment was done within cells based first on students' final school nonresponse cell and their explicit stratum; within that, grade and gender were used. Grade and gender were collected for all sampled students on the Student Tracking Form. Trimming factors at the school and student levels were also used to reduce the size of large weights, since large weights can substantially increase sampling variance. (One school weight was trimmed for the U.S. data; no student weights were trimmed.) All PISA analyses were conducted using these adjusted sampling weights. For more information on the nonresponse adjustments and trimming factors, see the OECD's *PISA 2006 Technical Report* (OECD 2009).

# Chapter 10.
# The PISA 2006 Data

## 10.1 PISA 2006 International Datasets

Data from PISA 2006 for all countries can be obtained from the OECD. At the time of this report's printing (April 2009), these data were available from http://pisa2006.acer.edu.au/. Users can either select to download entire files, choose only selected variables, or run simple queries. Files available for downloading include the following (note that the parent questionnaire and the information communication technology questionnaire items were not administered in the U.S.):

### Questionnaires

- Student questionnaire
- School questionnaire
- Parent questionnaire
- Information communication technology (ICT) questionnaire

### Codebooks

- Codebook for student questionnaire data file
- Codebook for school questionnaire data file
- Codebook for parent questionnaire data file
- Codebook for cognitive (assessment) and attitude item response data file
- Codebook for scored cognitive (assessment) and attitude item response data file

### SAS control files

- SAS syntax to read in student questionnaire data file
- SAS syntax to read in school questionnaire data file
- SAS syntax to read in parent questionnaire data file
- SAS syntax to read in cognitive (assessment) and attitude item response data file
- SAS syntax to read in scored cognitive (assessment) and attitude item response data file

### SPSS control files

- SPSS syntax to read in student questionnaire data file
- SPSS syntax to read in school questionnaire data file
- SPSS syntax to read in parent questionnaire data file
- SPSS syntax to read in cognitive (assessment) and attitude item response data file

- SPSS syntax to read in scored cognitive (assessment) and attitude item response data file

### Data sets in TXT format (compressed)

Note that some of these files are very large.

- Student questionnaire data file
- School questionnaire data file
- Parent questionnaire data file
- Cognitive (assessment) and attitude item response data file
- Scored cognitive (assessment) and attitude item response data file

### Compendia

The compendia provide the distribution of students according to the variables collected through the student, ICT, parent, and school questionnaires. The performance means per category are also provided.

- Compendium for the student questionnaire
- Compendium for the school questionnaire
- Compendium for the parent questionnaire
- Compendium for the ICT questionnaire
- Compendium for the cognitive item responses (Word)
- Compendium for the cognitive item responses (Excel)
- Compendium for the attitude item responses (Word)
- Compendium for the attitude item responses (Excel)

## 10.2 U.S. National Data Files

Data collected in the United States for PISA 2006 can be downloaded from the international site or from the NCES website. The files on the international website contain data for all countries, including the United States. The NCES files, which include only data for the United States, are as follows:

- Student data
  - The data are contained in STUD06.DAT. This file contains questionnaire items and derived variables and index scores based on the student questionnaire; plausible values for overall science, science competency subscales, science knowledge subscales, science attitude subscales, and the mathematics scale from the assessment; and student sampling weights and replicate weights. There are 5,611 cases in this file. Since the data are hierarchical (students are clustered with schools), each student record contains identification variables that enable the user to merge the school data with the student data, using the variable SCHOOLID.

  - an SPSS syntax file, STUD06.SPS
  - a SAS syntax file, STUD06.SAS
  - a codebook file (STUD06.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and frequencies.

- School data
  - The data are contained in SCHL06.DAT. This file contains items from the school questionnaire, derived variables and index scores based on the school questionnaire, and the school sampling weight. There are 166 cases in this file.
  - an SPSS syntax file, SCHL06.SPS
  - a SAS syntax file, SCHL06.SAS
  - a codebook file (SCHL06.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and frequencies.
- Cognitive (assessment) item data
  - The data are contained in ASSESM06.DAT. This file contains student responses to each item in the assessment. Note that the majority of the items have not been released, so there is little descriptive information about them. There are 5,611 cases in this file.
  - an SPSS syntax file, ASSESM06.SPS
  - a SAS syntax file, ASSESM06.SAS
  - a codebook file (ASSESM06.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and frequencies.
- U.S. Questionnaires
  - The U.S. version of the student questionnaire is in the file PISA_MS06_StudentQ_USA_Eng.PDF.
  - The U.S. version of the school questionnaire is in the file PISA_MS06_SchoolQ_USA_Eng.PDF
- Released items used in PISA 2006
  - Science items that were administered in PISA 2006 and subsequently released can be found in PISA_Sample_Items.PDF. Users looking for samples of mathematics and reading items should look at the items released after PISA 2003; all mathematics and reading items used in 2006 are still secure.
- PISA 2006 data user's guide
  - This document, PISA_2006_Data_Analysis_Users_Guide.PDF, contains information on the conduct of PISA in the United States.
- Macro for use with SPSS to produce plausible values and design-corrected standard errors

– This file, Setup_SPSSreplicates_V4_1.MSI, is designed for use with SPSS 15. It can be used to correct estimates of plausible values and design-corrected standard errors. A complete description of the macro can be found in *PISA 2003 Data Analysis Manual* (OECD 2005).

## 10.3 National and International Variables

The U.S. national data contain both the "international variables" (questionnaire and assessment variables used by all countries) and a few "national variables" (questionnaire variables used only in the United States). Note that the same assessment items were used by all countries. There are also some variables that appear in the international files that are missing for U.S. cases. Variables used only in the United States and those not used in the United States are listed here:

**Variables used only in the United States**

Student questionnaire

| Questionnaire item number | Variable name | Variable label |
|---|---|---|
| Q3a-Hispanic | RACE | Race – derived. Students in the United States were asked first whether they were Hispanic or Latino and then whether they were members of one or more of the following racial groups: White, Black, Asian, American Indian or Alaska Native, or Native Hawaiian/Other Pacific Islander. The recoded categories are: White (non-Hispanic), Black (non-Hispanic), Asian (non-Hispanic), Hispanic, Multi-racial non-Hispanic, and Other. |
| Q3b-White | | |
| Q3b-Black | | |
| Q3b-Asian | | |
| Q3b-American-Indian | | |
| Q3b-Pacific | | |
| Q5a | STQ5N01 | Mother currently doing |
| Q8a | STQ8N01 | Father currently doing |
| Q14e | ST14Q05 | How many rooms with bath or shower |
| Q30b | ST30N01 | Highest grade level - self |
| School questionnaire Q5 | SC05N01 | Percentage of students receiving free- or reduced price lunch |

**Variables not used in the United States**

Student questionnaire

- ST01Q02: Which of the following programs are you in?
- ST23Q02: ISCED Level 3B or 3C?
- ICT variables: These variables on the use of information technology were developed as a special add-on option for countries. The United States elected not to include these in the student questionnaires.

- PVREAD1 to PVREAD5: As explained in section 6.2, a printing error caused the U.S. results in reading to be declared invalid by the OECD and ACER. Consequently, plausible values in reading are not available for U.S. students.

School questionnaire:

- SC04Q13: Does school contain grade 13?

## 10.4 Confidentiality

The PISA 2006 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 and NCES 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 risks 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.

# Chapter 11. Using the PISA 2006 Data Files

## 11.1 Special Considerations—Plausible Values and Replicate Weights

Three aspects of the design of PISA need careful attention in any analysis. The first stems from the sample design. Schools and students had unequal but known probabilities of selection. As a consequence, to generalize to the population sampled, analyses will need to apply the sampling weights provided in the file. A detailed description of the procedures used in developing the weights for PISA is provided in the *PISA 2006 Technical Report* (OECD 2009).

The second aspect to be considered also stems from the sampling design and bears on the calculation of standard errors. Since the sample design is complex (a two-stage, stratified cluster design), most software packages, operating on the assumption of a simple random sample, will produce biased estimates of standard errors. Special procedures that use the replicate weights contained in the data file are called for, and are described in detail in the *PISA 2003 Data Analysis Manual* (OECD 2005, chapters 3 and 6; note that a PISA 2006 data analysis manual is being developed but is not currently available). These procedures are implemented in several stand-alone software packages (WesVar, AM, and SUDAAN, for example), but can also be implemented in SPSS using the macro posted on the NCES website, or in SAS using the information provided in chapter 15 in the *PISA 2003 Data Analysis Manual*. Standard errors produced in published reports were estimated using Fay's method of Balanced Repeated Replicates (BRR). That method should be specified when using SUDAAN or other stand-alone software packages to analyze the PISA data.

The third aspect arises from the design of PISA and the use of plausible values in analysis. In PISA, as in many national or international assessments, students are not administered every assessment item. Each item then has missing student responses, though these are missing by design. As a consequence, it is not possible to estimate scores for individual students. Instead, the results of individual students are aggregated to produce scores for groups of students (e.g., all U.S. students, U.S. female students, etc.). For analysis purposes, PISA datasets include sets of five "plausible values" for each student for each overall subject area score and each subscale score. The plausible values are intended to represent the estimated distribution of scores of students similar to the given student in terms of responses to the assessment items and background questionnaire items. What this means for analyses is that, in effect, any analyses involving the achievement scores must be done five times, once for each plausible value, and then the results must be averaged. A special provision also needs to be made in the estimation of the standard errors and is best done using the SPSS or SAS macro developed for this purpose. Again, these issues are discussed in the *PISA 2003 Data Analysis Manual* (OECD 2005, chapters 5 and 7).

## 11.2 Merging School and Student Data

The PISA sample was designed to yield a representative sample of 15-year-old students enrolled in schools; the school sample was designed to optimize the selection of these students. In these circumstances, it is usually recommended that the school data should be disaggregated across students and school attributes be treated as "student characteristics" for the purposes of the analyses (OECD 2005, chapter 9). This disaggregation can be accomplished by merging the school-level data

to the student file using SCHOOLID and the resulting file analyzed at the student level using the student-level weight (W_FSTUWT) or the replicate weights (W_FSTR1–W_FSTR80).

Merging school and student data is relatively easy, given the simple two-level structure of the data. Sample SPSS and SAS code that can be used to merge the files is presented in figures 1 and 2.

**Figure 1. Example of SPSS syntax for merging school and student data for PISA 2006**

```
get file 'c:\pisa\schl.sav'.
string subnatio (a4) .
compute subnatio=concat (country, subnat).
sort cases by subnatio schoolid .
save outfile='c:\pisa\schl.sav' .
```

```
get file='c:\pisa\stud.sav'.
sort cases by subnatio schoolid.
match files file=* /table='c:\pisa\schl.sav'
 /by subnatio schoolid .
save outfile='c:\pisa\merge.sav'.
```

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA) 2006.

**Figure 2. Example of SAS syntax for merging school and student data for PISA 2006**

```
data temp1
    set pisa2006.stud_US;
run;proc sort data = temp1
   By schoolid stidstd;
run;
data temp2;
   set pisa2006.schl_US;
run;
proc sort data = temp2;
   by schoolid;
run;
data pisa2006.alldata_US;
   merge temp1 temp2;
   by schoolid;
run;
```

Source: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA) 2006.

SPSS uses the file containing the school variables (schl.sav) and concatenates the file using the string variable subnatio and then sorts the cases by subnatio and school ID (schoolid). The file is then saved. The same procedures are used for the student dataset, stud.sav. The "match files" command merges the two files, and the final merged output is saved as 'merge.sav'.

The SAS example creates a temporary SAS dataset (temp1) using the permanent dataset (stud_usa.sas7bdat) (the "pisa2006" prefix is shorthand for the file's location on the user's computer). It then sorts the student data by school ID (schoolid) and student ID (stidstd). A similar procedure is used for the school file (temp2), which is sorted by schoolid. The final dataset will be a permanent dataset called alldata_usa.sas7bdat that contains the merged file using schoolid as the merge variable.

# References


Organization for Economic Cooperation and Development (OECD). (2005). *PISA 2003 Data Analysis Manual*. Paris: Author. Retrieved April, 14, 2008, from http://www.pisa.oecd.org/document/18/0,3343,en_32252351_32236173_35016146_1_1_1_1,00.html

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. (2009). *PISA 2006 Technical Report*. Paris: Organization for Economic Cooperation and Development.

Organization for Economic Cooperation and Development (OECD). (2007). *PISA 2006*: *Science Competencies for Tomorrow's World. Volume 1: Analysis*. Paris: Author. Retrieved April 14, 2008, from http://www.oecd.org/dataoecd/30/17/39703267.pdf

PISA Project Consortium. (2004). *Technical Standards for PISA 2006*. Retrieved April, 14, 2008, from https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2006_technical_standards.pdf

PISA Project Consortium. (2005a). *Main Study National Project Manager's Manual*. Retrieved April 14, 2008, from https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2003_national_project_manager_manual.pdf

PISA Project Consortium. (2005b). *School Sampling Preparation Manual: PISA 2006 Main Study*. Retrieved April 14, 2008, from https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2003_school_sampling_manual.pdf

PISA Project Consortium. (2005c). *PISA 2006 Main Study Test Administrator's Manual*. Retrieved April 14, 2008, from https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2006_test_administrator_manual.pdf

PISA Project Consortium. (2005d). *PISA 2006 Main Study School Coordinator's Manual*. Retrieved April 14, 2008, from https://mypisa.acer.edu.au/images/mypisadoc/opmanual/pisa2006_school_coordinators_manual.pdf

# Appendix A
# Comparing PISA with NAEP in Mathematics and Science

# Comparing PISA with NAEP in Mathematics and Science

Maria Stephens, American Institutes for Research

## A.1 Background

Both the Program for International Student Assessment (PISA) and the National Assessment of Educational Progress (NAEP, also known as "The Nation's Report Card") are primary sources for representative data on student achievement in the United States.[1] However, U.S. results from PISA are shaped by goals and standards of multiple participating countries, making interpretation from a U.S. perspective more challenging than the U.S.-designed and developed NAEP. The purpose of this appendix is to provide background information that will be useful in interpreting the results from PISA and to compare results from PISA with recent findings from similar subjects assessed in NAEP.

NAEP measures 4th-, 8th-, and 12th-grade students' performance in reading, mathematics, and science, with assessments designed specifically for national and state information needs. Alternatively, PISA enables the United States to benchmark its performance in 15-year-old students' reading, mathematics, and science literacy to that of other countries.[2] Both assessments are conducted regularly to allow the monitoring of student outcomes over time.[3]

Although PISA and NAEP assess similar content areas, PISA and NAEP were designed to serve different purposes and each is based on a separate and unique framework and set of assessment items (or questions). Thus, not surprisingly, there may be differences in results for a given year or in trend estimates between the studies, each giving a slightly different view of U.S. students' performance in these subjects.

This appendix is intended to provide information that will contribute to the understanding of results across studies, grasp the similarities and differences in these results, and identify what each assessment contributes to the overall knowledge base on student performance. To do so, it uses information from the 2006 administration of PISA and from the 2007 results from NAEP for 8th-grade mathematics and the 2005 NAEP results for 8th- and 12th-grade science and 12th-grade mathematics (see Baldi et al. 2007; Grigg, Donahue, and Dion 2007; Grigg, Lauko, and Brockway 2006; Lee, Grigg, and Dion 2007; Lee, Grigg, and Donahue 2007). Table A-1 summarizes the studies compared in this chapter.

---

[1] PISA is sponsored by the Organization for Economic Cooperation and Development (OECD). NAEP is sponsored by the National Center for Education Statistics, Institute of Education Sciences, in the U.S. Department of Education. The United States also participates in Progress in International Reading Literacy Study (PIRLS) and the Trends in International Mathematics and Science Study (TIMSS), both conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). See Neidorf, Binkley, and Stephens (2006) and Neidorf et al. (2006) for a comparison of TIMSS 2003 and NAEP.

[2] The 2006 PISA reading literacy assessment is not included in the comparisons in this paper. Because of a formatting error in the testing booklets and the small number of reading literacy items in the 2006 administration, the scores could not be recalibrated to exclude the affected items. Therefore, PISA 2006 reading literacy results are not reported for the United States.

[3] All statements about NAEP in this paper refer to national main NAEP (versus long-term trend NAEP). NAEP currently assesses 4th- and 8th-grade reading and mathematics every 2 years, and 12th-grade reading and mathematics, as well as science at all three grades, every 4 years. PISA is on a 3-year cycle.

**Table A-1. Scope of comparisons reported**

| Subject | 8th grade | 15-year-olds | 12th grade |
|---|---|---|---|
| Mathematics | NAEP 2007 | PISA 2006 | NAEP 2005 |
| Science | NAEP 2005 | PISA 2006 | NAEP 2005 |

NOTE: PISA is the Program in International Student Assessment. NAEP is the National Assessment of Educational Progress.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Program for International Student Assessment (PISA) 2006 and National Assessment of Educational Progress (NAEP) 2005 and 2007.

## A.2 Comparing Features of the Assessments

PISA and NAEP differ on several key features, including purpose, population, precision of estimates, and content.

### A.2.1 Purpose and Proximity to Curriculum

*The goals of PISA and NAEP have subtle but important distinctions with regard to the U.S. curricula.*

Using nationally established benchmarks of performance (e.g., basic, proficient, advanced), NAEP is the U.S. source for information on reading, mathematics, and science achievement at key stages of education across the country. The frameworks and benchmarks are established by the National Assessment Governing Board (NAGB) and are based on the collaborative input of a wide range of experts and participants from government, education, business and public sectors in the United States. Ultimately, they are intended to reflect the best thinking about the knowledge, skills, and competencies needed by U.S. students to have an in-depth understanding of these subjects at different grades.

PISA is the U.S. source for internationally comparative information on the reading, mathematics, and science literacy of students in the upper grades at an age that, for most countries, is near the end of compulsory schooling. The objective of PISA is to measure the "yield" of education systems, or what skills and competencies students have acquired and can apply in these subjects to real-world contexts by age 15. The literacy concept emphasizes the mastery of processes, understanding of concepts, and application of knowledge and functioning in various situations within domains. By focusing on literacy, PISA draws not only from school curricula but also from learning that may occur outside of school.

The tailoring of NAEP to national practices distinguishes it from PISA, the content of which is determined internationally in collaboration with other countries and which reflects a consensus view of key content. The focus in PISA on yield and the application of competencies in real-world contexts is distinct from NAEP's focus on measuring school-based curricular attainment more closely.

### A.2.2 Population

*PISA provides benchmarks associated with a diverse group of countries.*

NAEP is designed to produce estimates of student achievement in the United States while PISA is designed to provide comparable assessment results across a range of developed and developing countries. The Organization for Economic Cooperation and Development (OECD) sponsors PISA, with its 30 member countries representing the world's most industrialized nations. A

total of 57 countries participated in the 2006 administration, including all 30 OECD countries (table A-2). However, the international average in PISA is based only on the OECD countries' scores.

**Table A-2. Country participation in PISA (2006)**

| Country group | Participating countries | |
|---|---|---|
| OECD countries | Australia | Luxembourg |
| | Austria | Mexico |
| | Belgium | Netherlands |
| | Canada | New Zealand |
| | Czech Republic | Norway |
| | Denmark | Poland |
| | Finland | Portugal |
| | France | Slovak Republic |
| | Germany | Spain |
| | Greece | Sweden |
| | Hungary | Switzerland |
| | Iceland | Turkey |
| | Ireland | United Kingdom |
| | Italy | United States |
| | Japan | |
| | Korea (South) | |
| Other countries | Argentina | Kyrgyz Republic |
| | Azerbaijan | Latvia |
| | Brazil | Lithuania |
| | Bulgaria | Macao-China |
| | Chile | Qatar |
| | Chinese Taipei | Republic of Montenegro |
| | Colombia | Republic of Serbia |
| | Croatia | Russian Federation |
| | Estonia | Romania |
| | Hong Kong | Slovenia |
| | Indonesia | Thailand |
| | Israel | Tunisia |
| | Jordan | Uruguay |

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

*The students being studied may represent different groups.*

NAEP and PISA are sample-based assessments—meaning that each program administers the assessment to a subgroup of U.S. students in such a way that the results can be generalized to the larger population. However, each assessment defines the population to which it is generalizing (and thus from which the sample is drawn) differently. For example, PISA uses an age-based sample of 15-year-olds, while NAEP uses a grade-based sample. These choices relate to the purposes of each program described earlier—NAEP to report on curricular achievement and PISA to describe the yield of systems toward the end of compulsory schooling.

- The NAEP target population is all students in 4th, 8th, and 12th grades, and thus reflects the performance of U.S. students in these grades—most recently for 4th- and 8th-grade reading and mathematics in 2007, 12th-grade reading and mathematics in 2005, and all three grades in science in 2005.

- The PISA target population is all 15-year-old students. Operationally in 2006, this included all students who were from 15 years and 3 months to 16 years and 2 months at the beginning of the testing period and who were enrolled in school, regardless of grade level or full- or part-time status. The PISA results reflect the performance of U.S. 15-year-old students, who were mostly in 10th grade, and the rest in other grades.

In the upper grades, the PISA population is uniformly older than NAEP 8th-graders and uniformly younger than the NAEP 12th-graders. NAEP and PISA also are assessing different cohorts in different years. Taking this into account, perhaps the closest NAEP-PISA comparisons can be made between the NAEP 2005 8th-grade and PISA 2006 15- year-old student cohorts, some of the former of whom theoretically could have been part of the latter. However, all side-by-side comparisons of NAEP and PISA results should be viewed with these population and cohort differences in mind.

### A.2.3 Precision of Estimates

*The assessments are designed to measure at different levels of precision.*

NAEP and PISA are designed to provide valid and reliable measures of U.S. students' performance in the aggregate and for major subpopulations, and each study draws a sample sufficient for this purpose. NAEP and PISA differ, however, in the size of the differences in performance they are intended to detect. Student performance varies widely across countries and so PISA is designed to detect relatively large differences. NAEP is designed to detect smaller differences. NAEP can detect smaller variations in student performance within the U.S. states than across the many countries participating in PISA, as well as smaller variations in performance over time. It is important for NAEP to be sensitive to small changes in student performance over time, for the nation as a whole, and for individual states.

Sample sizes are calculated to balance needs for precision of estimates against burden to respondents. Because of NAEP's need for greater precision, NAEP samples have many more students than PISA within the U.S. Table A-3 shows the sample size for NAEP and the U.S. sample size for PISA.

**Table A-3. Sample sizes in NAEP and PISA in the United States**

| Study | Number of students sampled | Number of schools sampled |
|---|---|---|
| NAEP 2007 (8th grade) | 160,700 | 6,930 |
| PISA 2006 | 5,611 | 166 |

SOURCE: U.S. Department of Education, National Center for Education Statistics, Program for International Student Assessment (PISA) 2006 and National Assessment of Educational Progress (NAEP) 2007.

### A.2.4 Content

*The mathematics and science material being assessed may be different in terms of the ways in which the frameworks for assessment are organized and in terms of content coverage, item format, and other key features.*

As noted before, the assessments under discussion here are developed from frameworks that define the domain and specify the content and skills to be measured. Thus, a first task in comparing assessment programs is to compare how the frameworks and specifications are elaborated. A second

task, which can provide a more in-depth view is to compare how the frameworks are operationalized through the actual assessment items.

### A.2.4.1 Science: PISA 2006 and NAEP 2005

There has not been an extensive study comparing how PISA and NAEP each define and assess science for the upper grades. However, we can look to the respective frameworks for some insight as to similarities and differences. By necessity, we examine NAEP's framework for the 2005 assessment, even though the next administration in 2009 will be grounded in a revised science framework. The PISA framework examined here is that which was elaborated for the most recent assessment of science literacy in 2006.

In these documents, both PISA and NAEP emphasize that science extends beyond knowledge of scientific facts to include broad understanding of science concepts and knowledge of how to apply and use scientific concepts and skills. There is recognition in both documents that it is important for students to demonstrate knowledge about science itself—what distinguishes science from other ways of knowing (or understanding what it can and cannot answer), how to approach issues and evaluate information scientifically, and what is its role and impact and interaction with man and society. However, each framework is organized differently and therefore, until there is a more comprehensive study of items, it will be difficult to say how similar or different the actual assessments might be.

The science framework for NAEP 2005 is organized around a matrix, with content and cognitive dimensions, as well as two overarching dimensions. The content dimension includes the fields of science (life, physical, and Earth sciences) in which students demonstrate their cognitive skills of conceptual understanding, scientific investigation, and practical reasoning. The overarching dimensions—meaning that a certain number of items within the assessment will also meet additional characteristics—include items on the nature of science (such as understanding the scientific process, the interaction of man with the world, the role of technology in science) and on interdisciplinary themes, such as the idea of scientific models, the notion of systems, and patterns of change. The NAEP framework also designates a number of items to be "hands-on" tasks involving the use of materials to conduct scientific investigations, which represent an item format unique to the NAEP assessment and one relevant to science instruction in the United States.

The PISA framework also has content and cognitive dimensions, although as an immediate organizational difference, it seems that they are broader, capturing some of what likely is included in NAEP's separately titled overarching dimensions. PISA's content dimension, for example, includes both knowledge of science (in the fields of life systems, physical systems, Earth and space systems, and technology systems) and knowledge about science (scientific inquiry and scientific explanations). PISA's competencies also are specified somewhat differently and seem to be centered around an explicit dissection of scientific inquiry: identifying scientific issues, explaining scientific phenomena, and using scientific evidence. In the PISA model, the competencies are prominent—they form the subscales for reporting, for instance—and are shown as influenced by the content/knowledge dimension. The PISA framework also is explicit about the situationally based nature of science literacy and thus has a context dimension that describes a range of situations in which individuals deploy their competencies. It also puts an attitudinal dimension alongside the content/knowledge dimension and embeds attitudinal items (normally placed in a background questionnaire) within the actual assessment. Like students' knowledge, their attitudes (interest and motivation, sense of responsibility, and support for inquiry) are seen in the PISA model as influences on competencies.

Although the responses to the attitudinal items are not part of the PISA score—they are reported separately—their presence in the assessment is a feature unique to PISA.

Given how differently the frameworks are organized, comparing intended item distributions is difficult. Looking at the content/knowledge dimension, NAEP calls for a roughly even distribution of items across the fields of science (physical, life, Earth), with a slightly heavier emphasis on life science at the 8th-grade level. PISA, on the other hand, divides its knowledge dimension into more categories and thus the three fields of science it has in common with NAEP each represents no more than a quarter of the items. While NAEP items may have additional classifications (such as being a "nature of science" or "themes" item), each one ties back to a specific scientific discipline, whereas the PISA framework suggests that it has some items that are solely on knowledge *about* science rather than knowledge *of* science. Looking at the cognitive dimension, NAEP emphasizes conceptual understanding, with nearly half of all items meeting this definition. PISA aims for a slightly more even distribution across its competency clusters. But again, because the categories are quite different, comparisons are limited.

The additional elements of the PISA science framework and other apparent differences likely reflect its overall purpose to study the application of knowledge and skills in real-world contexts in which individuals are interacting and participating in society. On the other hand, the organization of the NAEP framework and its relatively more content-oriented nature is by definition grounded in assessing the school-based learning of students in the context of U.S. standards and instruction.

### A.2.4.2 Mathematics: PISA 2006 and NAEP 2005/7

At first glance, there are some noticeable similarities in the structure of the NAEP and PISA mathematics frameworks, as summarized in an earlier comparison study.[4] For instance—although differently titled—both NAEP and PISA are organized (primarily) along a content dimension and a cognitive dimension. In NAEP, these are the content strands and levels of mathematical complexity, respectively, and in PISA, these are the overarching ideas and competency clusters. However, the manner in which the content and cognitive dimensions are further specified within these dimensions differs between the two—reflecting NAEP's close ties to the organizational structures used in traditional school curricula and, by contrast, PISA's focus on the application of mathematics in real-world situations. Thus, the PISA framework also includes a third dimension, which is the situation or context (e.g., educational or personal) of an item.

NAEP's content strands include five major areas of mathematics, including number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions—within which specific topics and subtopics are further identified. In PISA, the content areas are described in terms of overarching ideas, which include change and relationships; quantity; space and shape; and uncertainty. NAEP's levels of mathematical complexity (denoted as low, moderate and high) identify where along a continuum of cognitive demand an item falls, with skills like the ability to perform a stated procedure at one end and the ability to engage in abstract reasoning or generalize a pattern at the other end. PISA's competency clusters, while not an explicit or exact hierarchy, similarly denote three sets of gradually increasing cognitive demand, from reproduction to connection to reflection.

[4] The information in this section is taken from the technical report on a comparison study that was undertaken for the NAEP 2003 and PISA 2003 mathematics assessments. See Neidorf et al. (2006). This source is used because there have been no changes to the frameworks (that were not already accounted for in the study) between the 2003 and most recent administrations of the two assessments.

These structural and terminological differences suggest that there may be differences in the ways in which NAEP and PISA are operationalized and, indeed, the previous comparison study of NAEP and PISA mathematics items has shed light on the degree to which this is true.[5] In terms of content similarities, the past study showed that there is a considerable overlap between PISA's uncertainty overarching idea and NAEP's data analysis, statistics, and probability content strand, as well as the space and shape idea and the geometry and spatial sense strands. In terms of content differences, however, PISA was shown to have a relatively greater focus on data analysis, statistics, and probability and lesser focus on algebra than the NAEP 8th-grade assessment.[6]

In terms of the cognitive dimensions, PISA had more items requiring reflection, which includes less familiar or more complex problem settings and a higher demand for thinking and reasoning and developing and communicating an argument, than did NAEP at both 8th- and 12th-grade. NAEP, on the other hand, had more items that fell into the reproduction category, which includes reproducing practiced material and performing routine operations.[7] However, while the past study concluded that PISA items classified in a higher level of cognitive complexity and demand than NAEP items, the content covered was most consistent with the topics specified in the NAEP 8th-grade framework.

Finally, it is important to note the different distributions of items of a particular format in the two programs—with NAEP aiming for a roughly even split between multiple-choice and constructed response (i.e., open-ended) items and PISA aiming for a greater emphasis on constructed response, with about two-thirds of items of such a nature. While item type is not necessarily directly related to item difficulty, students' ease or familiarity with different item types may systematically differ and thus contribute to differences in results.

## A.3 Results in the Context of Assessment Differences

Both PISA and NAEP provide measures of mathematics and science performance for older students (in grades 8 to 12). It is natural to compare them, but the distinctions described previously need to be kept in mind in understanding converging or diverging results.

### A.3.1 Comparing Select Results for Science

It is difficult to compare the results from PISA and NAEP in science, not only for the population and framework differences, but also because PISA is not yet reporting a trend measure for science, which would be the most likely element to examine in the context of NAEP's trend measure. The last assessment of science in NAEP (2005), showed no statistically significant differences in the performance of 8th- or 12th-graders since 2000 (although there was a slight decrease since 1996 among the older students).

---

[5] It should be noted that the percentages reported in the footnotes for this section are for the sets of items used in the 2003 administrations of NAEP and PISA; item-level comparisons were not redone with the NAEP 2005/7 and PISA 2006 items. Because the PISA 2006 items are a subset of the 2003 items and because the newer NAEP items will include some items repeated from 2005/07, as well as newer items designed to replicate the items being replaced, the general conclusions remain valid. The percentages should be taken as illustrative rather than exact, however.

[6] Thirty-nine percent of PISA items were classified to the NAEP data category, compared with 10 percent of 8th-grade and 25 percent of 12th-grade items in NAEP. In contrast, 9 percent of PISA items were classified as algebra, compared with 15 percent of 8th-grade and 35 percent of 12th-grade items in NAEP.

[7] Thirty-one percent of PISA items were classified in the reproduction competency cluster, compared with 58 and 39 percent of NAEP 8th- and 12th-grade problem-solving items, respectively. On the other hand, 22 percent of PISA's items were deemed to fit the reflection competency cluster, compared to 5 percent or less of NAEP items at both grades.

### A.3.2 Comparing Select Results for Mathematics

In mathematics at the upper grades, PISA 2006 shows that statistically there is no change in the scores of U.S. 15-year-olds since 2003. On PISA 2006, the U.S. score for mathematics literacy is below the average for all OECD countries. NAEP shows an increase in the scores of 8th-grade mathematics students between 2003 and 2007. Again, NAEP's design allows it to pick up small changes in the performance of U.S. students.

## A.4 Summary

In sum, there appears to be an advantage in capitalizing on the complementary information presented in national and international assessments. NAEP measures in detail the reading, mathematics, and science knowledge of U.S. students as a whole, and can also provide trend information for individual states, different geographic regions, and demographic population groups. International assessments like PISA add value by providing a method for comparing performance in the United States to the performance of students in other nations. However, differences in study design and content need to be recognized when interpreting results. Some of the differences between NAEP and PISA include:

- The goals of the assessments have subtle but important distinctions with regard to the U.S. curricula. NAEP is tailored specifically to practices and standards operating in the United States, which distinguishes it from PISA, the content of which is determined internationally in collaboration with other countries and reflecting consensus views of key content. Also, PISA's specific focus on the yield of the education system and the application of competencies in real-word contexts, distinguishes it from NAEP, which aims at measuring school-based curricular attainment more closely.
- The students being studied represent different groups. NAEP uses grade-based samples and targets 4th-, 8th-, and 12th-grade students. PISA uses an age-based sample, which targets 15-year-olds, who likely are between the NAEP target populations of 8th- and 12th-graders.
- The assessments are designed to measure student performance at different levels of precision. NAEP and PISA are designed to provide valid and reliable measures of U.S. students' performance in the aggregate and for major subpopulations, and each study draws a sample sufficient for this purpose. NAEP, however, is designed to also provide estimates for individual states, which requires an increased sample size and thus produces performance measures at a higher level of precision than PISA. These differences can have an impact on the assessments' sensitivities in detecting changes in student performance.
- The mathematics and science being assessed can be different in terms of the ways in which the frameworks for assessment are organized and in terms of content coverage, item format, and other key features. Examinations of the frameworks for NAEP and PISA in reading, mathematics, and science show areas of potential overlap and potential difference in terms of the content and skills being measured in the respective subject areas and grades.

## A.5 References

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

Grigg, .W., Donahue, P.L., and Dion, G. (2007). *Nation's Report Card: 12th-Grade Reading and Mathematics 2005* (NCES 2007-468). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Grigg, W., Lauko, M.A., and Brockway, D.M. (2006). *Nation's Report Card: Science 2005* (NCES 2006-466). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Lee, J., Grigg, W.S., and Dion, G.S. (2007). *The Nation's Report Card: Mathematics 2007* (NCES 2007-494). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Lee, J., Grigg, W.S., and Donahue, P.L. (2007). *The Nation's Report Card: Reading 2007* (NCES 2007-496). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Neidorf, T., Binkley, M., and Stephens, M. (2006). *Comparing Science Content in the National Assessment of Educational Progress (NAEP) 2000 and Trends in International Mathematics and Science Study (TIMSS) 2003 Assessments* (NCES 2006-026). Washington, DC: National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education.

Neidorf, T., Binkley, M., Gattis, K., and Nohara, D. (2006). *Comparing Mathematics Content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments* (NCES 2006-029). Washington, DC: National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education.

# Appendix B
# Nonresponse Bias Analysis

# B.1 Introduction

## B.1.1 Response Rates and Purpose of Nonresponse Bias Analysis

An important component of survey or assessment data quality is the representativeness of the study sample. This representativeness of the population is achieved by selecting a sample of respondents who are similar to the population in terms of key demographic markers. However, in practice not all sampled respondents participate in surveys. If enough respondents fail to participate or if respondents differ in their response rates by key demographic characteristics, the final sample may not represent the target population. The extent to which the distributions of the sampled respondents differ from the corresponding distributions of the population is termed unit nonresponse bias. One way to characterize and quantify the presence of unit nonresponse bias is to compare responding sample elements with nonresponding sample elements with respect to underlying sociodemographic characteristics for which data are available on the frame. Also, not all sample respondents respond to all applicable items in the questionnaire. If enough respondents fail to respond to a question or if respondents differ in their item response rates by key demographic characteristics, the item respondents may not represent the student respondents. The extent to which the distributions of the item respondents differ from the corresponding distributions of the sample respondents is termed item nonresponse bias. One way to characterize and quantify the presence of item nonresponse bias is to compare responding item elements with nonresponding item elements with respect to underlying sociodemographic characteristics for which data are available on the frame. The National Center for Education Statistics (NCES) requires a unit nonresponse bias analysis for all datasets based on surveys in which the unit response rate is less than 85 percent and an item nonresponse bias analysis for all questionnaire items for which the response rate is less than 85 percent. This report presents the results of unit and item nonresponse bias analyses for the Program for International Student Assessment (PISA) 2006 U.S. data collection, which had a school response rate of 69 percent and some items with response rates below the NCES 85 percent threshold. The report also presents results of an analysis studying differences between schools in which the test was administered during school hours and those that opted to have the test administered outside of school hours, either after school or on a Saturday.

The objective of this nonresponse bias analysis is to shed light on any biases that might be present in the data because of nonresponse. That is, we analyze responding and nonresponding schools to determine whether responding schools are representative of the original sample or whether there are significant differences between the responding and nonresponding schools. Also, we analyze respondents and nonrespondents to questionnaire items to determine whether students responding to items actually represent a virtually random subsample of the sample respondents or whether there are significant differences between the students responding and not responding to items.

The second objective of this report is to study the differences between schools in which the test was administered during school hours and the remaining schools in which the test was administered at another time. This feature of the survey was introduced to increase the willingness of schools to participate but the question arises as to whether this might have introduced response bias into the final data.

## B.1.2 Overview of PISA

PISA is sponsored by the Organization for Economic Cooperation and Development and funded in the United States by NCES, Institute of Education Sciences, U.S. Department of Education. The assessment is designed to measure the cumulative knowledge of students at age 15. Thus, the U.S. sample was drawn to be representative of the population of 15-year-old students in the United States. Fifteen-year-old students were defined as all students born between July 1, 1990, and June 30, 1991, who were enrolled in grades 7–12.

### B.1.2.1 Sampling

The sample was drawn in two stages. Schools were sampled in the first stage of the sampling process using probabilities proportional to size. In the second stage students were sampled in all selected schools that agreed to participate. The target cluster size for sampling students within schools was 42 students.

Within each responding country, a minimum of 150 schools were selected for PISA 2006. In the United States, the PISA sample was designed to be representative of all 15-year-old students in the 50 states and the District of Columbia.

The U.S. school sample for PISA 2006 was drawn in March 2005. The sampling frame was constructed using data from the 2003–04 Common Core of Data (CCD) and preliminary data from the 2003–04 Private School Universe Survey (PSS). Before the selection process, schools were sorted into two explicit groups: schools with large enrollments of 15-year-old students and schools with small enrollments of 15-year old students. The schools were then sorted by implicit strata: grade span of school, percentage of racial/ethnic minority students, control of school (public/private), percentage of students eligible for free or reduced-priced lunch, and locale. Schools were selected on the basis of the number of 15-year-old students in the school, so that schools with more students in 9th and 10th grades had a higher probability of selection than schools with fewer grades containing 15-year-olds. The final sample included 236 schools.

In accordance with PISA guidelines, at the same time as the PISA sample was selected, substitute (replacement) schools were identified by assigning the two schools neighboring the sampled school in the frame as substitutes. Substitutes were required to be in the same implicit stratum as the sampled school. If the sampled school was the first or last school in the stratum, then the second school following or preceding the sampled school was identified as the substitute. One was designated a first substitute and the other a second substitute. If an original school refused to participate, the first substitute was then contacted. If that school also refused to participate, the second school was then contacted.

PISA was conducted in the fall of 2006 in the United States, beginning on September 25 and ending on November 22. Of the 236 original sample schools, 209 were found to be eligible (18 schools did not have any 15-year-olds enrolled, 5 had closed, and 4 were alternative schools for behavioral issues where students returned to a base school after a short period of time), and 145 agreed to participate. The weighted and unweighted school response rate before substitution was 69 percent. In addition to the 145 participating original schools, 21 substitute schools participated for a total of 166 participating schools, or 79 percent overall response rate.[1] The PISA standards call for a

[1] 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 substitute schools that participated in the denominator as well as the numerator, and to add substitute schools that were hard refusals to the denominator. This results in a response rate of 67.5 percent.

minimum participation rate of 65 percent of original sample schools, a standard which the United States met. However, since the U.S. response rate level did not meet the NCES standard of 85 percent, the nonresponse bias analysis was conducted to shed light on the quality of the data.

### B.1.2.2 Statistical Analysis

All schools were initially assigned a basic (design) weight that is the inverse of the probability that the school would be selected for the sample. A school-level participation (nonresponse) weight adjustment was then made to compensate for any sampled schools that did not participate and were not substituted. This adjustment was done by the international consortium and was calculated independently for each explicit stratum. The resulting weight after adjustment is referred to in this report as the weight after (nonresponse) adjustment and also as the final nonresponse adjusted weight.

The unit nonresponse bias analysis consists of a comparison of the characteristics of the respondents with those of the nonrespondents before weight adjustment and a comparison of the characteristics of the respondents with those of the full sample after weight adjustment to determine whether the respondents represent in any way a biased subsample of the original sample. Throughout the report, the tables that were produced using the design weight before adjustment compare the responding schools with the nonresponding schools. In these tables, the bias is computed as the estimated mean or percentage of the responding schools subtracted from the estimated mean or percentage of the nonresponding schools, and the relative bias is computed as the bias divided by the estimated mean or percentage of the responding schools. The relative bias is a ratio or percentage, and when the bias is the difference of two percentages, the bias is a percentage. The tables that are based on final nonresponse adjusted weights compare responding schools with the full sample. The respondents cannot be compared with the nonrespondents after weight adjustment because the respondent weights have been adjusted upward to represent the nonrespondents, and the nonrespondents are then dropped from the analysis file. The respondents are compared with the full sample to see if the final weighted distributions are similar to the original sample distributions. Creating a data file containing both the responding sample and full sample appended together and using SUDAAN allows for comparisons of these dependent samples. In these tables, the bias is computed as the estimated mean or percentage of the responding schools subtracted from the estimated mean or percentage of the full sample, and the relative bias is computed as the bias divided by the estimated mean or percentage of the responding schools.

The next section implements this analysis for the original sample of 209 eligible schools attempting to identify any bias in the group of 145 schools that originally responded. Section B.3 repeats the analysis for the same-sized sample of 209 schools but this time, there are 21 substitute schools included as responding schools. In each section, the analysis first studies categorical variables using a chi-square statistic to test differences between responding and nonresponding schools before weight adjustment and between responding and full sample schools after weight adjustment. The second part of the analysis focuses on continuous variables and uses the *t*-statistic to test differences in means between respondents and nonrespondents before weight adjustment and between respondents and the full sample after weight adjustment. The variables used in the analysis are from the NCES CCD and PSS. All analyses were conducted using SUDAAN. As an additional effort to identify any presence of bias, we employ logistic regression analysis with response status as the dependent variable.

Typically, in most NCES studies, the nonresponse bias analysis is conducted as two steps in the construction of the weight variables. After design weights are constructed, the nonresponse bias

analysis is conducted, and a nonresponse adjustment to the weight variable is made based on the results. Then, the nonresponse bias analysis is conducted a second time after nonresponse adjustment to see how much bias remains. In the international studies, the weight construction task is carried out by the international consortium, and standard procedures are used for all participating countries. Nonresponse adjustment to the school weight is based only on the explicit strata used to select the sample (see the technical notes in section B.9 for further information about the weighting). Nonresponse bias analysis was not used to inform the development of the weights. Therefore, the purpose of the nonresponse bias analysis is to determine whether the nonresponse adjustment changed the bias observed using the design weight and to determine how much bias remains in the data that analysts need to be aware of when making inferences from the population.

The item nonresponse bias analysis consists of a comparison of the characteristics of the item respondents and the nonrespondents to determine whether the respondents represent in any way a biased subsample of the student respondents. Section B.7 implements this analysis for the school and student questionnaire items with a response rate less than 85 percent. Questionnaire items are analyzed by categorical variables, some of which are continuous variables categorized based on quartiles. For each category of each variable, bias is computed as the percentage of all item respondents who are in that category minus the percentage of all item nonrespondents who are in that category. The *t*-statistic was used to test the significance of this bias for each category of each variable. The school-level and student-level analyses were both conducted using weights. The variables used in the item-level analysis are also from the NCES CCD and PSS. All item analyses were also conducted using SUDAAN.

To investigate the effect of test administration during school hours compared with after school or Saturday administrations, we analyze the two groups of schools with respect to a number of underlying characteristics and to test scores. The comparisons are carried out both in a bivariate model comparing means between the two groups and in a multivariate context in which the test scores are modeled as dependent variables with a number of explanatory independent variables in the model.

Most of the analyses both with respect to nonresponse bias and the impact of test administration were carried out at the school level. Given the clustered nature of the sample, it is expected that the majority of the variance will occur between schools rather than within schools. Any statistically valid inference that holds at the school level will most likely also apply at the student level.

## B.2. Nonresponse Bias Analysis: Original Sample

This section presents the results of the unit nonresponse bias analysis based on the original sample of 209 eligible schools. All schools that were substituted were treated as nonrespondents, as were any nonresponding original schools. There were 145 schools that originally responded and 64 schools that did not respond. Standard errors are given throughout. The first analysis compares the distribution of the responding schools with that of the nonresponding schools using weighted data in each case for three categorical and three continuous variables. The weights are based on the design weight, which is the inverse of the probability of selection of the school. The sample of schools was drawn using probabilities proportional to size where the size measure is the total number of students enrolled in fourth grade on the sampling frame. The second analysis compares the distribution of the responding schools with the eligible (full sample) schools. The weights for this second analysis are based on the final weight which was adjusted for nonresponse for the

responding schools and the design weight for the eligible schools. The nonresponse weight adjustment is described in section B.9.

### B.2.1 Categorical Variables—Before Nonresponse Weight Adjustment

The following variables were available for both public and private schools:

- public/private affiliation;
- community type; and
- Census region; the regions are defined in section B.9.

**Public/Private Affiliation.** Table B-1 shows the school distribution by public and private affiliation for respondents and nonrespondents. The test of independence indicates that there is no evidence of a significant difference between response status and public/private affiliation at the 5 percent level. However, the absolute values of the relative bias for both public and private schools are greater than 10 percent, which indicates potential bias even though no statistically significant difference was detected.

**Table B-1. Percentage distribution of original sample schools, for respondents and nonrespondents before nonresponse weight adjustment, by school control: Fall 2006**

| School characteristic | Respondents | | Nonrespondents | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| School control | | | | | | | 0.276 |
| Public | 77.2 | 8.98 | 90.1 | 6.67 | 12.92 | 16.74 | |
| Private | 22.8 | 8.98 | 9.9 | 6.67 | -12.92 | -56.67 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Community Type.** Table B-2 shows school distribution by community type for respondents and nonrespondents. The test of independence indicates that there is no evidence of a significant difference between response status and community type at the 5 percent level. However, the absolute values of the relative bias for central city and urban fringe or large town schools are greater than 10 percent, which indicates potential bias even though no statistically significant difference was detected.

**Table B-2. Percentage distribution of original sample schools, for respondents and nonrespondents before nonresponse weight adjustment, by community type: Fall 2006**

| School characteristic | Respondents | | Nonrespondents | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| Community type | | | | | | | 0.536 |
| Central city | 25.6 | 8.73 | 16.1 | 6.12 | -9.43 | -36.89 | |
| Urban fringe or large town | 22.2 | 7.21 | 29.7 | 9.28 | 7.56 | 34.08 | |
| Rural or small town | 52.3 | 9.46 | 54.1 | 13.27 | 1.87 | 3.58 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Census Region.** Table B-3 shows school distribution by census region for respondents and nonrespondents. The test of independence indicates that there is no evidence of a significant difference between response status and Census region at the 5 percent level. However, the absolute values of the relative bias for all four regions are greater than 10 percent, which indicates potential bias even though no statistically significant difference was detected.

**Table B-3. Percentage distribution of original sample schools, for respondents and nonrespondents before nonresponse weight adjustment, by Census region: Fall 2006**

| School characteristic | Respondents | | Nonrespondents | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| Census region | | | | | | | 0.593 |
| Northeast | 11.3 | 2.99 | 14.5 | 5.17 | 3.17 | 27.98 | |
| South | 35.4 | 8.33 | 27.1 | 9.12 | -8.35 | 23.56 | |
| Midwest | 37.0 | 8.93 | 22.6 | 8.76 | -14.37 | 38.86 | |
| West | 16.2 | 6.64 | 35.8 | 16.61 | 19.56 | 120.37 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample. See section B.9 for state listings within regions.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

### B.2.2 Categorical Variables—After Nonresponse Weight Adjustment

**Public/Private Affiliation.** Table B-4 shows school distribution by public and private affiliation for respondents and the full sample. The test of independence indicates that there is no evidence of a significant difference between response status and public/private affiliation at the 5 percent level. Compared with the bias before weight adjustment, the absolute values of the bias after weight adjustment decreased. Also, the absolute values of the relative bias for both public and private schools are less than 10 percent.

**Table B-4. Percentage distribution of original sample schools, for respondents and the full sample after nonresponse weight adjustment, by school control: Fall 2006**

| School characteristic | Respondents | | Full sample | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| School control | | | | | | | 0.510 |
| Public | 78.4 | 8.81 | 80.1 | 7.29 | 1.66 | 2.12 | |
| Private | 21.6 | 8.81 | 19.9 | 7.29 | -1.66 | -7.70 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Community Type.** Table B-5 shows school distribution by community type for respondents and the full sample. The test of independence indicates that there is no evidence of a significant difference between response status and community type at the 5 percent level. Compared with the bias before weight adjustment, the absolute values of the bias after weight adjustment decreased for two of the three community types. Also, only the absolute value of the relative bias for central city schools is greater than 10 percent.

**Table B-5. Percentage distribution of original sample schools, for respondents and the full sample after nonresponse weight adjustment, by community type: Fall 2006**

| School characteristic | Respondents | | Full sample | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| Community type | | | | | | | 0.428 |
| Central city | 27.3 | 8.97 | 23.5 | 7.03 | -3.81 | -13.97 | |
| Urban fringe or large town | 23.4 | 7.13 | 23.9 | 5.90 | 0.46 | 1.96 | |
| Rural or small town | 49.3 | 9.35 | 52.7 | 7.95 | 3.36 | 6.81 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Census Region.** Table B-6 shows school distribution by census region for respondents and the full sample. The test of independence indicates that there is no evidence of a significant difference between response status and Census region at the 5 percent level. Compared with the bias before weight adjustment, the absolute values of the bias after weight adjustment decreased for all four regions. Also, the absolute values of the relative bias for all regions, except the West, are less than 10 percent.

**Table B-6. Percentage distribution of original sample schools, for respondents and the full sample after nonresponse weight adjustment, by Census region: Fall 2006**

| School characteristic | Respondents | | Full sample | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| Census region | | | | | | | 0.784 |
| Northeast | 12.0 | 2.56 | 12.0 | 2.29 | 0.01 | # | |
| South | 34.7 | 8.03 | 33.6 | 6.61 | -0.15 | 3.31 | |
| Midwest | 36.8 | 8.84 | 33.8 | 7.15 | -3.00 | 8.16 | |
| West | 16.5 | 6.35 | 20.6 | 6.60 | 4.15 | 25.20 | |

# Rounds to zero.

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample. See section B.9 for state listings within regions.

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

### B.2.3 Continuous Variables—Before Nonresponse Weight Adjustment

The following variables were available for both public and private schools:

- total number of students;
- number of age-eligible students enrolled;
- percentage of Asian /Pacific Islander, non-Hispanic students;
- percentage of Black, non-Hispanic students;
- percentage of Hispanic students;
- percentage of American Indian or Alaska Native, non-Hispanic students;
- percentage of White, non-Hispanic students;
- percentage of other students; and
- percentage of students eligible for free or reduced-price lunch.

**Age-Eligible Enrollment and Total Students.** Table B-7 shows the school-level mean total and age-eligible number of students for the respondents and nonrespondents. There is no evidence of a significant difference between the mean total or age-eligible number of students of responding and nonresponding schools. However, the absolute values of the relative bias for both the mean total and age-eligible number of students are greater than 10 percent, which indicates potential bias even though no statistically significant differences were detected.

**Table B-7. Mean age-eligible enrollment and total students in original sample schools, for respondents and nonrespondents before nonresponse weight adjustment: Fall 2006**

| School characteristic | Respondents | | Nonrespondents | | | | |
|---|---|---|---|---|---|---|---|
| | Total | Standard error | Total | Standard error | Bias | Relative bias | *t* test *p*-value |
| Total number of students | 531.7 | 73.25 | 752.7 | 180.18 | 221.06 | 40.58 | 0.257 |
| Age-eligible students | 114.5 | 19.67 | 179.3 | 45.45 | 64.77 | 56.56 | 0.192 |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Race/Ethnicity Enrollment.** Table B-8 presents results based on race/ethnicity enrollment for original sample schools. The table presents mean percentage enrollment in each race/ ethnicity category averaged over all schools for the respondents and nonrespondents. There is evidence of a significant difference between the mean percentage enrollment of responding and nonresponding schools only for the Other race/ethnicity category. However, the absolute values of the relative bias for the other five categories are greater than 10 percent, which indicates potential bias even though no statistically significant differences were detected.

**Table B-8. Percentage distribution of student enrollment in original sample schools for respondents and nonrespondents before nonresponse weight adjustment, by race/ethnicity: Fall 2006**

| Race/ethnicity | Respondents | | Nonrespondents | | | | |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | Bias | Relative bias | *t* test *p*-value |
| Asian/Pacific Islander, non-Hispanic | 4.1 | 0.66 | 5.7 | 1.16 | 1.61 | 39.46 | 0.247 |
| Black, non-Hispanic | 18.2 | 5.67 | 12.6 | 2.30 | -5.62 | 30.81 | 0.367 |
| Hispanic | 11.1 | 1.60 | 9.4 | 1.69 | -1.64 | -14.81 | 0.494 |
| American Indian/Alaska Native, non-Hispanic | 1.7 | 0.83 | 3.2 | 2.52 | 1.51 | 88.82 | 0.567 |
| White, non-Hispanic | 61.7 | 4.76 | 68.8 | 3.68 | 7.16 | 11.61 | 0.245 |
| Other | 3.2 | 1.29 | 0.2 | 0.11 | -3.02 | -93.21 | 0.020 |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample. 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.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Eligibility for Free or Reduced-Price Lunch.** Table B-9 shows the mean percentage of free or reduced-price lunch eligible students averaged over all schools for the respondents and nonrespondents. There is no evidence of a significant difference between the mean percentage of free or reduced-price lunch eligible students of responding and nonresponding schools. However, the absolute value of the relative bias is greater than 10 percent, which indicates potential bias even though no statistically significant difference was detected.

**Table B-9. Percentage distribution of students eligible for free or reduced-price lunch in original sample public schools, for respondents and nonrespondents before nonresponse weight adjustment: Fall 2006**

| Students | Respondents | | Nonrespondents | | Bias | Relative bias | *t* test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| Students eligible for free or reduced-price lunch | 28.3 | 4.73 | 23.0 | 2.91 | -5.35 | -18.87 | 0.341 |

NOTE: Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

## B.2.4 Continuous Variables—After Nonresponse Weight Adjustment

**Age-Eligible Enrollment and Total Students.** Table B-10 shows the school-level mean total and age-eligible number of students for the respondents and full sample. There is no evidence of a significant difference between the mean total or age-eligible number of students of responding and full sample schools. Compared with the bias before weight adjustment, the absolute values of the bias after weight adjustment decreased. Also, the absolute values of the relative bias for the mean total and age-eligible number of students are less than 10 percent.

**Table B-10. Mean age-eligible enrollment and total students in original sample schools, for respondents and the full sample after nonresponse weight adjustment: Fall 2006**

| School characteristic | Respondents | | Full sample | | Bias | Relative bias | *t* test *p*-value |
|---|---|---|---|---|---|---|---|
| | Total | Standard error | Total | Standard error | | | |
| Total number of students | 562.3 | 75.85 | 581.1 | 70.24 | 18.74 | 3.33 | 0.581 |
| Age-eligible students | 122.6 | 21.07 | 129.0 | 18.73 | 6.44 | 5.25 | 0.452 |

NOTE: Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Race/Ethnicity Enrollment.** Table B-11 presents results based on race/ethnicity enrollment for original sample schools. The table presents mean percentage enrollment in each race/ ethnicity category averaged over all schools for the respondents and full sample. There is evidence of a significant difference between the mean percentage enrollment of responding and full sample schools only for the Other race/ethnicity category. Compared with the bias before weight adjustment, the absolute values of the bias after weight adjustment decreased for all race/ethnicity categories. Also, the absolute values of the relative bias for race/ethnicity are greater than 10 percent for the Black and American Indian/Alaska Native categories.

**Table B-11. Percentage distribution of student enrollment in original sample schools for respondents and the full sample after nonresponse weight adjustment, by race/ethnicity: Fall 2006**

| Race/ethnicity | Respondents | | Full sample | | | | |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | Bias | Relative bias | *t* test *p*-value |
| Asian/Pacific Islander, non-Hispanic | 4.2 | 0.67 | 4.5 | 0.56 | 0.31 | 7.31 | 0.455 |
| Black, non-Hispanic | 18.9 | 5.95 | 16.6 | 4.12 | -2.26 | -11.97 | 0.268 |
| Hispanic | 11.6 | 1.71 | 10.6 | 1.20 | -1.01 | -8.70 | 0.215 |
| American Indian/ Alaska Native, non-Hispanic | 1.6 | 0.71 | 2.1 | 0.94 | 0.59 | 38.06 | 0.432 |
| White, non-Hispanic | 60.6 | 4.90 | 63.7 | 3.58 | 3.13 | 5.17 | 0.103 |
| Other | 3.1 | 1.25 | 2.4 | 0.92 | -0.75 | -24.04 | 0.048 |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample. 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.

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

**Eligibility for Free or Reduced-Price Lunch.** Table B-12 shows the mean percentage of free or reduced-price lunch eligible students averaged over all schools for the respondents and full sample. There is no evidence of a significant difference between the mean percentage of free or reduced-price lunch eligible students of responding and full sample schools. Compared with the bias before weight adjustment, the absolute value of the bias after weight adjustment decreased. Also, the absolute value of the relative bias for the mean percentage of free or reduced-price lunch eligible students is less than 10 percent.

**Table B-12. Percentage distribution of students eligible for free or reduced-price lunch in original sample public schools, for respondents and the full sample after nonresponse weight adjustment: Fall 2006**

| Students | Respondents | | Full sample | | | | |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | Bias | Relative bias | *t* test *p*-value |
| Students eligible for free or reduced-price lunch | 28.3 | 4.90 | 26.8 | 3.50 | -1.47 | -5.20 | 0.408 |

NOTE: Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.

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

## B.2.5 Logistic Regression Model

We implemented a stepwise logistic regression to identify underlying factors that are related to response propensity. The stepwise selection method in SAS "proc logistic" indicated that all variables included in this initial model were significant, and these variables were then modeled using SUDAAN to account for the complex sample design with the design weight. In SUDAAN, the dependent response variable was coded as '0' for nonrespondents and '1' for respondents. In addition to the intercept, the school characteristics significantly different between responding and

nonresponding schools were the percentage of Black, American Indian, and Asian students, which had higher estimates for nonresponding schools than for responding schools. The result for Black students is inconsistent with the results in table B-8 due to the race/ethnicity variable being defined differently[2]. The SUDAAN estimates, standard errors, test statistics, and *p*-values are reported in table B-13.

**Table B-13. Final model parameters for original sample schools: Fall 2006**

| Parameter | Estimate | Standard error | Test for $H_0$: parameter = 0 | *p*-value |
|---|---|---|---|---|
| Intercept | 41.75 | 19.86 | 2.10 | 0.004 |
| Publically controlled school | -0.28 | 1.08 | -0.26 | 0.792 |
| Northeast | 0.62 | 0.70 | 0.88 | 0.379 |
| South | 0.49 | 0.77 | 0.64 | 0.525 |
| Midwest | 0.34 | 0.65 | 0.53 | 0.596 |
| Central city | -0.20 | 0.93 | -0.21 | 0.831 |
| Urban fringe or large town | -0.67 | 0.77 | -0.87 | 0.385 |
| Age-eligible students | 0.00 | 0.01 | 0.53 | 0.598 |
| Total school enrollment | -0.00 | 0.00 | -0.83 | 0.407 |
| Students eligible for free or reduced-price lunch | 0.03 | 0.02 | 1.34 | 0.180 |
| Percent Hispanic | -0.39 | 0.21 | -1.87 | 0.062 |
| Percent White, non-Hispanic | -0.40 | 0.21 | -1.97 | 0.051 |
| Percent Black, non-Hispanic | -0.41 | 0.21 | -1.98 | 0.049 |
| Percent American Indian/Alaska Native, non-Hispanic | -0.46 | 0.21 | -2.18 | 0.030 |
| Percent Asian/Pacific Islander, non-Hispanic | -0.41 | 0.20 | -2.01 | 0.045 |

NOTE: $H_0$ is the null hypothesis that the parameter is equal to 0. See section B.9 for state listings within regions. 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.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

## B.3 Nonresponse Bias Analysis: Sample with Substitutes (Final Sample)

The response rate of original schools, 69.4 percent, was sufficiently high to meet the PISA standard of a minimum of 65 percent. Substitute schools were recruited to increase the sample size allowing for more in-depth analysis and tighter confidence intervals. The school response rate including substitute schools was 79.4 percent. Following is a nonresponse bias analysis on the final sample of schools, including substitute schools. The nonresponse bias analysis is based exclusively on the final sample of 209 eligible schools, and nonresponding schools are any nonparticipating schools that were not replaced by a substitute school. The first analysis compares the distribution of the 166 responding schools with that of the 43 nonresponding schools using weighted data in each case for three categorical and three continuous variables. The weights are based on the design weight, which is the inverse of the probability of selection of the school. The sample of schools was drawn using probabilities proportional to size where the size measure is the total number of students

[2] For table B-8, race/ethnicity was defined as the percentage distribution of student enrollment across all responding schools and across all nonresponding schools. For table B-13, the percentage for each race/ethnicity category was based on each school. The other race category was created and used in table B-8 so that the total percentages sum to 100 percent. However, the other race category was not used in table B-13 because it was not on the CCD and PSS files and was not considered an important variable for the purpose of the model.

enrolled in fourth grade on the sampling frame. The second analysis compares the distribution of the responding schools with the eligible (full sample) schools. The weights for this second analysis are based on the final weight adjusted for nonresponse for the responding schools and the design weight for the eligible schools. The nonresponse weight adjustment is described in section B.9.

### B.3.1 Categorical Variables—Before Nonresponse Weight Adjustment

The following characteristics were available for both public and private schools:

- public/private affiliation;
- community type; and
- Census region.

**Public/Private Affiliation.** Table B-14 shows school distribution by public and private affiliation for respondents and nonrespondents. The test of independence indicates that there is no evidence of a significant difference between response status and public/private affiliation at the 5 percent level. However, the absolute values of the relative bias for both public and private schools are greater than 10 percent, which indicates potential bias even though no statistically significant difference was detected.

**Table B-14. Percentage distribution of final sample schools, for respondents and nonrespondents, by school control before nonresponse weight adjustment: Fall 2006**

| School characteristic | Respondents | | Nonrespondents | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| School control | | | | | | | 0.117 |
| Public | 77.1 | 8.25 | 95.7 | 4.45 | 18.61 | 24.13 | |
| Private | 22.9 | 8.25 | 4.3 | 4.45 | -18.61 | -81.30 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Community Type.** Table B-15 shows school distribution by community type for respondents and nonrespondents. The test of independence indicates that there is no evidence of a significant difference between response status and community type at the 5 percent level. However, the absolute values of the relative bias for all three community types are greater than 10 percent, which indicates potential bias even though no statistically significant difference was detected.

**Table B-15. Percentage distribution of final sample schools, for respondents and nonrespondents, by community type before nonresponse weight adjustment: Fall 2006**

| School characteristic | Respondents | | Nonrespondents | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| Community type | | | | | | | 0.327 |
| Central city | 24.1 | 7.95 | 15.5 | 8.14 | -8.57 | -35.60 | |
| Urban fringe or large town | 22.5 | 6.60 | 37.3 | 15.11 | 14.82 | 65.93 | |
| Rural or small town | 53.5 | 8.66 | 47.2 | 20.39 | -6.24 | -11.67 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Census Region.** Table B-16 shows school distribution by Census region for respondents and nonrespondents. The test of independence indicates that there is no evidence of a significant difference between response status and Census region at the 5 percent level. However, the absolute values of the relative bias for all four regions are greater than 10 percent, which indicates potential bias even though no statistically significant difference was detected.

**Table B-16. Percentage distribution of final sample schools, for respondents and nonrespondents, by Census region before nonresponse weight adjustment: Fall 2006**

| School characteristic | Respondents | | Nonrespondents | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| Census region | | | | | | | 0.180 |
| Northeast | 10.8 | 2.63 | 17.9 | 8.17 | 7.10 | 65.68 | |
| South | 36.8 | 7.52 | 18.0 | 8.65 | -18.86 | -51.24 | |
| Midwest | 37.1 | 8.05 | 14.6 | 6.91 | -22.49 | -60.62 | |
| West | 25.3 | 5.99 | 49.5 | 19.29 | 24.27 | 96.04 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample. See section B.9 for state listings within regions.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

### B.3.2 Categorical Variables—After Nonresponse Weight Adjustment

**Public/Private Affiliation.** Table B-17 shows school distribution by public and private affiliation for respondents and the full sample. The test of independence indicates that there is no evidence of a significant difference between response status and public/private affiliation at the 5 percent level. Compared with the bias before weight adjustment, the absolute values of the bias after weight adjustment decreased. Also, the absolute values of the relative bias for both public and private schools are less than 10 percent.

**Table B-17. Percentage distribution of final sample schools, for respondents and the full sample after nonresponse weight adjustment, by school control: Fall 2006**

| School characteristic | Respondents | | Full sample | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| School control | | | | | | | 0.399 |
| Public | 78.5 | 8.08 | 79.9 | 7.22 | 1.43 | 1.82 | |
| Private | 21.5 | 8.08 | 20.1 | 7.22 | -1.43 | -6.64 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Community Type.** Table B-18 shows school distribution by community type for respondents and the full sample. The test of independence indicates that there is no evidence of a significant difference between response status and community type at the 5 percent level. Compared with the bias before weight adjustment, the absolute values of the bias after weight adjustment decreased for all three community types. Also, only the absolute value of the relative bias for central city schools is greater than 10 percent.

**Table B-18. Percentage distribution of final sample schools, for respondents and the full sample after nonresponse weight adjustment, by community type: Fall 2006**

| School characteristic | Respondents | | Full sample | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| Community type | | | | | | | 0.353 |
| Central city | 25.8 | 8.19 | 22.8 | 6.96 | -3.01 | -11.67 | |
| Urban fringe or large town | 24.0 | 6.53 | 24.7 | 5.90 | 0.75 | 3.13 | |
| Rural or small town | 50.3 | 8.55 | 52.5 | 7.89 | 2.27 | 4.52 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Census Region.** Table B-19 shows school distribution by Census region for respondents and the full sample. The test of independence indicates that there is no evidence of a significant difference between response status and Census region at the 5 percent level. Compared with the bias before weight adjustment, the absolute values of the bias after weight adjustment decreased for all four regions. Also, the absolute values of the relative bias for all regions, except the West, are less than 10 percent.

**Table B-19. Percentage distribution of final sample schools, for respondents and the full sample after nonresponse weight adjustment, by Census region: Fall 2006**

| School characteristic | Respondents | | Full sample | | Bias | Relative bias | Chi-square test *p*-value |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | | | |
| Census region | | | | | | | 0.538 |
| Northeast | 11.7 | 2.32 | 11.9 | 2.28 | 0.21 | 1.80 | |
| South | 35.9 | 7.30 | 34.0 | 6.55 | -1.88 | -5.24 | |
| Midwest | 36.9 | 7.64 | 33.7 | 7.07 | -3.18 | -8.62 | |
| West | 15.6 | 5.76 | 20.4 | 6.54 | 4.84 | 31.11 | |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample. See section B.9 for state listings within regions.

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

### B.3.3 Continuous Variables—Before Nonresponse Weight Adjustment

The following variables were available for both public and private schools:

- total number of students;
- number of age-eligible students enrolled;
- percentage of Asian or Pacific Islander, non-Hispanic students;
- percentage of Black, non-Hispanic students;
- percentage of Hispanic students;
- percentage of American Indian or Alaska Native, non-Hispanic students;
- percentage of White, non-Hispanic students;
- percentage of Other students; and
- percentage of students eligible for free or reduced-price lunch.

**Age-Eligible Enrollment and Total Students.** Table B-20 shows the school-level mean total and age-eligible number of students for the respondents and nonrespondents. There is no evidence of a significant difference between the mean total or age-eligible number of students of responding and nonresponding schools. However, the absolute values of the relative bias for both the mean total and age-eligible number of students are greater than 10 percent, which indicates potential bias even though no statistically significant differences were detected.

**Table B-20. Mean age-eligible enrollment and total students in final sample schools, for respondents and nonrespondents before nonresponse weight adjustment: Fall 2006**

| School characteristic | Respondents | | Nonrespondents | | | | |
|---|---|---|---|---|---|---|---|
| | Total | Standard error | Total | Standard error | Bias | Relative bias | *t* test *p*-value |
| Total number of students | 548.6 | 68.90 | 736.7 | 250.02 | 188.12 | 34.29 | 0.468 |
| Age-eligible students | 118.7 | 18.56 | 178.3 | 63.00 | 59.60 | 50.20 | 0.364 |

NOTE: Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Race/Ethnicity Enrollment.** Table B-21 presents results based on race/ethnicity enrollment for final sample schools. The table presents mean percentage enrollment in each race/ethnicity category averaged over all schools for the respondents and nonrespondents. There is evidence of a significant difference between the mean percentage enrollment of responding and nonresponding schools only for the Other race/ethnicity category. However, the absolute values of the relative bias for the other five categories are greater than 10 percent, which indicates potential bias even though no statistically significant differences were detected.

**Table B-21. Percentage distribution of student enrollment in original sample schools for respondents and nonrespondents before nonresponse weight adjustment, by race/ethnicity: Fall 2006**

| Race/ethnicity | Respondents | | Nonrespondents | | | | |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | Bias | Relative bias | *t* test *p*-value |
| Asian/Pacific Islander, non-Hispanic | 4.0 | 0.59 | 4.6 | 0.90 | 0.58 | 14.43 | 0.599 |
| Black, non-Hispanic | 17.5 | 5.03 | 14.4 | 3.13 | -3.07 | -17.57 | 0.612 |
| Hispanic | 11.2 | 1.47 | 8.9 | 1.81 | -2.35 | -20.91 | 0.330 |
| American Indian/Alaska Native, non-Hispanic | 1.6 | 0.73 | 4.7 | 3.75 | 3.10 | 198.72 | 0.419 |
| White, non-Hispanic | 62.7 | 4.31 | 67.1 | 4.43 | 4.40 | 70.16 | 0.486 |
| Other | 3.0 | 1.13 | 0.3 | 0.16 | -2.67 | -89.00 | 0.020 |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample. 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.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Eligibility for Free or Reduced-Price Lunch.** Table B-22 shows the mean percentage of free or reduced-price lunch eligible students averaged over all schools for the respondents and nonrespondents. There is no evidence of a significant difference between the mean percentage of free or reduced-price lunch eligible students of responding and nonresponding schools. However, the absolute value of the relative bias is greater than 10 percent, which indicates potential bias even though no statistically significant difference was detected.

**Table B-22. Percentage distribution of students eligible for free or reduced-price lunch in final sample public schools, for respondents and nonrespondents before nonresponse weight adjustment: Fall 2006**

| Students | Respondents | | Nonrespondents | | | | |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | Bias | Relative bias | *t* test *p*-value |
| Students eligible for free or reduced-price lunch | 28.2 | 4.21 | 22.3 | 3.78 | -5.96 | -21.12 | 0.298 |

NOTE: Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the nonresponding sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

## B.3.4 Continuous Variables—After Nonresponse Weight Adjustment

**Age-Eligible Enrollment and Total Students.** Table B-23 shows the school-level mean total and age-eligible number of students for the respondents and full sample. There is no evidence of a significant difference between the mean total or age-eligible number of students of responding and full sample schools. Compared with the bias before weight adjustment, the absolute values of the bias after weight adjustment decreased. Also, the absolute values of the relative bias for the mean total and age-eligible number of students are less than 10 percent.

**Table B-23. Mean age-eligible enrollment and total students in final sample schools, for respondents and the full sample after nonresponse weight adjustment: Fall 2006**

| School characteristic | Respondents | | Full sample | | | | |
|---|---|---|---|---|---|---|---|
| | Total | Standard error | Total | Standard error | Bias | Relative bias | *t* test *p*-value |
| Total number of students | 581.0 | 71.51 | 576.7 | 68.76 | -4.22 | -0.73 | 0.892 |
| Age-eligible students | 127.5 | 19.92 | 127.7 | 18.33 | 0.17 | 0.13 | 0.962 |

NOTE: Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Race/Ethnicity Enrollment.** Table B-24 presents results based on race/ethnicity enrollment for original sample schools. The table presents mean percentage enrollment in each race/ ethnicity category averaged over all schools for the respondents and full sample. There is evidence of a significant difference between the mean percentage enrollment of responding and full sample schools only for the Hispanic race/ethnicity category. Compared with the bias before weight adjustment, the absolute values of the bias after weight adjustment decreased for all race/ethnicity categories. Also, the absolute values of the relative bias for the mean total and age-eligible number of students are greater than 10 percent for the Other and American Indian/Alaska Native categories.

**Table B-24. Percentage distribution of student enrollment in original sample schools for respondents and the full sample after nonresponse weight adjustment, by race/ethnicity: Fall 2006**

| Race/ethnicity | Respondents | | Full sample | | | | |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | Bias | Relative bias | *t* test *p*-value |
| Asian/Pacific Islander, non-Hispanic | 4.2 | 0.61 | 4.1 | 0.50 | -0.09 | -2.13 | 0.716 |
| Black, non-Hispanic | 18.0 | 5.28 | 16.9 | 43.11 | -1.15 | -6.38 | 0.404 |
| Hispanic | 12.0 | 1.60 | 10.8 | 1.21 | -1.18 | -9.86 | 0.048 |
| American Indian/Alaska Native, non-Hispanic | 1.4 | 0.63 | 2.2 | 0.94 | 0.71 | 49.31 | 0.328 |
| White, non-Hispanic | 61.5 | 4.44 | 63.5 | 3.58 | 2.08 | 3.38 | 0.121 |
| Other | 2.9 | 1.10 | 2.5 | 0.92 | -0.38 | -13.24 | 0.104 |

NOTE: Detail may not sum to totals because of rounding. Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample. 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.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

**Eligibility for Free or Reduced-Price Lunch.** Table B-25 shows the mean percentage of free or reduced-price lunch eligible students averaged over all schools for the respondents and full sample. There is no evidence of a significant difference between the mean percentage of free or reduced-price lunch eligible students of responding and full sample schools. Compared with the bias before weight adjustment, the absolute value of the bias after weight adjustment decreased. Also, the absolute value of the relative bias for the mean percentage of free or reduced-price lunch eligible students is less than 10 percent.

**Table B-25. Percentage distribution of students eligible for free or reduced-price lunch in final sample public schools, for respondents and the full sample after nonresponse weight adjustment: Fall 2006**

| Students | Respondents | | Full sample | | | | |
|---|---|---|---|---|---|---|---|
| | Percent | Standard error | Percent | Standard error | Bias | Relative bias | *t* test *p*-value |
| Students eligible for free or reduced-price lunch | 28.2 | 4.35 | 27.1 | 3.50 | -1.08 | -3.84 | 0.380 |

NOTE: Nonresponding schools are eligible schools that did not agree to have their students assessed. Responding schools agreed to have their students assessed. Bias is calculated as the estimate from the full sample minus the estimate from the responding sample. The relative bias is calculated as the bias divided by the estimate from the responding sample.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

## B.3.5 Logistic Regression Model

We implemented a stepwise logistic regression to identify underlying factors that are related to response propensity. The stepwise selection method in SAS "proc logistic" indicated that all variables included in this initial model, except percent Hispanic, were significant, and these variables were then modeled using SUDAAN to account for the complex sample design with the design weight. In SUDAAN, the dependent response variable was coded as '0' for nonrespondents and '1' for respondents. In addition to the intercept, the school characteristics significantly different between responding and nonresponding schools were publicly controlled schools, urban fringe or

large town, and the percentage of White and American Indian/Alaska Native students, which all had higher estimates for nonresponding schools than for responding schools. Also, the South region had higher estimates for responding schools than for nonresponding schools. The SUDAAN estimates, standard errors, test statistics, and *p*-values are reported in table B-26.

**Table B-26. Final (reduced) model parameters for final sample schools: Fall 2006**

| Parameter | Estimate | Standard error | Test for $H_0$: parameter = 0 | *p*-value |
|---|---|---|---|---|
| Intercept | 9.03 | 1.79 | 5.03 | 0.000 |
| Publically controlled school | -2.52 | 1.19 | -2.11 | 0.036 |
| Northeast | 0.54 | 0.84 | 0.65 | 0.519 |
| South | 1.61 | 0.80 | 2.02 | 0.044 |
| Midwest | 1.16 | 0.76 | 1.53 | 0.127 |
| Central city | -1.82 | 1.23 | -1.48 | 0.140 |
| Urban fringe or large town | -2.57 | 0.96 | -2.66 | 0.008 |
| Age-eligible students | 0.00 | 0.01 | 0.03 | 0.977 |
| Total school enrollment | -0.00 | 0.00 | -0.24 | 0.809 |
| Students eligible for free or reduced-price lunch | 0.02 | 0.01 | 1.52 | 0.130 |
| Percent White, non-Hispanic | -0.04 | 0.02 | -2.00 | 0.047 |
| Percent Black, non-Hispanic | -0.05 | 0.03 | -1.83 | 0.069 |
| Percent American Indian/Alaska Native, non-Hispanic | -0.11 | 0.02 | -4.52 | 0.000 |
| Percent Asian/Pacific Islander, non-Hispanic | -0.04 | 0.05 | -0.82 | 0.415 |

NOTE: $H_0$ is the null hypothesis that the parameter is equal to 0. See section B.9 for state listings within regions. 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.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

The results in table B-27 indicate the results of the full model with percent Hispanic included. The school characteristics significantly different between responding and nonresponding schools were publicly controlled schools, urban fringe or large town, and the South region. The estimate for the publicly controlled schools and urban fringe or large town estimates was higher for nonresponding schools than for responding schools, and the estimate for the South region was higher for responding schools than for nonresponding schools. The SUDAAN estimates, standard errors, test statistics, and *p*-values are reported in table B-27 for the full model.

**Table B-27. Final (full) model parameters for final sample schools: Fall 2006**

| Parameter | Estimate | Standard error | Test for $H_0$: parameter = 0 | *p*-value |
|---|---|---|---|---|
| Intercept | 21.62 | 14.97 | 1.44 | 0.150 |
| Publically controlled school | -2.13 | 1.18 | -1.82 | 0.072 |
| Northeast | 0.61 | 0.85 | 0.71 | 0.478 |
| South | 1.57 | 0.77 | 2.04 | 0.042 |
| Midwest | 1.07 | 0.74 | 1.45 | 0.150 |
| Central city | -1.83 | 1.22 | -1.52 | 0.129 |
| Urban fringe or large town | -2.58 | 0.96 | -2.67 | 0.008 |
| Age-eligible students | 0.00 | 0.01 | 0.19 | 0.847 |
| Total school enrollment | -0.00 | 0.00 | -0.39 | 0.700 |
| Students eligible for free or reduced-price lunch | 0.02 | 0.01 | 1.64 | 0.103 |
| Percent Hispanic | -0.14 | 0.16 | -0.89 | 0.376 |
| Percent White, non-Hispanic | -0.17 | 0.16 | -1.10 | 0.271 |
| Percent Black, non-Hispanic | -0.18 | 0.16 | -1.16 | 0.249 |
| Percent American Indian/Alaska Native, non-Hispanic | -0.24 | 0.16 | -1.53 | 0.127 |
| Percent Asian/Pacific Islander, non-Hispanic | -0.17 | 0.17 | -1.00 | 0.319 |

NOTE: $H_0$ is the null hypothesis that the parameter is equal to 0. See section B.9 for state listings within regions. 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.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

## B.4 Test Administration and School Characteristics

The following sections of the report focus on the issue of test administration; specifically, whether schools differed with respect to the type of administration and whether there is a relationship between type of test administration and test scores. Schools were given the option of conducting the testing during the normal school day (during school hours) or after school or on a Saturday (outside of school hours). The first part of the analysis in this section compares the two groups of schools with respect to various school characteristics in an effort to determine whether there are any significant differences between the two groups. Where possible, we included the same variables that were used in the previous set of analyses carried out in this report. There were 88 schools with in-school test administration and 78 schools with out-of-school test administration. In section B.5, we compare the two groups of schools with respect to their average scores on the mathematics and science overall scales and science subscales using statistical confidence intervals for differences between the two group means. Finally, in section B.6, we subject the data to a multivariate regression model using score as the dependent variable and various underlying variables, including type of test administration, as the explanatory factors. Tests for differences in student scores were implemented at both the school and student levels, and no differences were found between the two groups of schools. In all cases, SUDAAN was used to test the statistical significance and data were weighted based on the original or final sample schools using the final weights.

Table B-28 presents weighted distributions comparing characteristics of final sample schools in which tests were administered in-school versus those in which the tests were administered out of school. The test administration bias (difference) and relative bias are also shown. *T* tests were conducted for each paired comparison, and no significant differences were found. However, the

absolute values of the relative bias for all but two of the characteristics are greater than 10 percent, which indicates potential bias even though no statistically significant differences were detected.

**Table B-28. A comparison of schools administering PISA during and outside of schools hours, by various school characteristics: Fall 2006**

| School characteristic | In-school | | Out-of-school | | Bias | Relative bias | *p*-value for *t* test |
|---|---|---|---|---|---|---|---|
| | Estimate | Standard error | Estimate | Standard error | | | |
| Public | 70.0 | 11.76 | 92.2 | 4.68 | 22.26 | 31.81 | 0.076 |
| Private | 30.3 | 11.76 | 7.8 | 4.68 | -22.56 | -74.38 | 0.076 |
| Central city | 22.2 | 9.48 | 31.4 | 14.29 | 9.20 | 41.42 | 0.593 |
| Urban fringe or large town | 22.2 | 9.49 | 26.7 | 8.26 | 4.52 | 20.37 | 0.720 |
| Rural or small town | 55.6 | 11.21 | 41.9 | 13.11 | -13.72 | -24.68 | 0.428 |
| Northeast | 6.7 | 2.66 | 19.4 | 6.14 | 12.62 | 187.00 | 0.089 |
| Southeast | 39.7 | 10.61 | 32.5 | 12.69 | -7.22 | -18.17 | 0.682 |
| Central | 35.2 | 9.25 | 36.9 | 13.49 | 1.16 | 4.57 | 0.926 |
| West | 18.3 | 9.27 | 11.3 | 3.60 | -7.01 | -38.33 | 0.504 |
| Total school enrollment | 491.3 | 82.48 | 721.2 | 125.48 | 229.95 | 46.81 | 0.130 |
| Age-eligible students | 109.2 | 21.54 | 156.1 | 39.63 | 46.85 | 42.90 | 0.302 |
| Percent Asian/Pacific Islander, non-Hispanic | 3.9 | 0.67 | 4.6 | 1.13 | 0.65 | 16.67 | 0.637 |
| Percent Black, non-Hispanic | 12.4 | 2.19 | 24.1 | 9.93 | 11.72 | 94.82 | 0.254 |
| Percent Hispanic | 10.8 | 1.99 | 13.2 | 2.69 | 2.46 | 22.82 | 0.471 |
| Percent American Indian/Alaska Native, non-Hispanic | 2.1 | 1.16 | 0.8 | 0.27 | -1.29 | -62.61 | 0.281 |
| Percent White, non-Hispanic | 66.6 | 3.28 | 56.1 | 7.93 | -10.50 | -15.78 | 0.228 |
| Percent Other | 4.3 | 1.82 | 1.3 | 1.11 | -3.04 | -70.05 | 0.158 |
| Students eligible for free or reduced-priced lunch | 478.3 | 109.80 | 443.6 | 101.88 | -34.63 | -7.24 | 0.230 |

NOTE: Bias is calculated as the estimate from the out-of-school test administration minus the estimate from the in-school test administration. The relative bias is calculated as the bias divided by the estimate from the in-school test administration. See section B.9 for state listings within regions. 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.

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

## B.5 Test Administration and Scores

To test the presence of a test administration effect on scores, schools with in-school administration and those with out-of-school administration were compared on their average test scores (including mathematics and science literacy and subscales). We also pooled all students in each condition (tested during school hours versus tested outside of school hours) and compared their average scores. Results are presented separately for the original sample and the final sample. The test administration bias (difference) and relative bias are also shown. Weighted data using final weights were used in all cases.

### B.5.1 School-Level Scores: Original Sample

Table B-29 presents results based on five score categories for original sample schools. The table presents estimates in each score category for the schools with in-school and out-of-school

administration. There is no evidence of a significant difference between the scores of schools administering the test during school hours and those administering it outside of school hours.

**Table B-29. Mean score in original sample schools, for administrations during school hours and outside of school hours, by score category: Fall 2006**

| Score category | In-school | | Out-of-school | | Bias | Relative bias | *p*-value for *t* test |
|---|---|---|---|---|---|---|---|
| | Estimate | Standard error | Estimate | Standard error | | | |
| Math | 486.1 | 8.12 | 445.7 | 23.98 | -40.34 | -8.30 | 0.115 |
| Science | 511.1 | 8.25 | 462.9 | 28.00 | -48.12 | -9.42 | 0.106 |
| Explaining phenomena scientifically | 509.6 | 9.10 | 458.2 | 25.68 | -51.40 | -10.09 | 0.064 |
| Identifying scientific issues | 509.6 | 8.12 | 480.5 | 26.18 | -29.10 | -5.71 | 0.294 |
| Using scientific evidence | 511.9 | 9.83 | 455.8 | 37.11 | -56.13 | -10.97 | 0.150 |

NOTE: Bias is calculated as the estimate from the out-of-school test administration minus the estimate from the in-school test administration. The relative bias is calculated as the bias divided by the estimate from the in-school test administration.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

## B.5.2 School-Level Scores: Final Sample

Table B-30 presents results based on five score categories for final sample schools. The table presents estimates in each score category for the schools with in-school and out-of-school administration. There is no evidence of a significant difference between the scores of schools administering the test during school hours and those administering it outside of school hours.

**Table B-30. Mean score in final sample schools, for administrations during and outside of school hours, by score category: Fall 2006**

| Score category | In-school | | Out-of-school | | Bias | Relative bias | *p*-value for *t* test |
|---|---|---|---|---|---|---|---|
| | Estimate | Standard error | Estimate | Standard error | | | |
| Math | 484.5 | 7.52 | 448.2 | 21.85 | -36.27 | -7.49 | 0.120 |
| Science | 509.3 | 7.95 | 465.8 | 25.42 | -43.55 | -8.55 | 0.108 |
| Explaining phenomena scientifically | 508.2 | 8.61 | 461.6 | 23.49 | -46.61 | -9.17 | 0.067 |
| Identifying scientific issues | 508.7 | 7.63 | 480.7 | 23.42 | -28.00 | -5.50 | 0.261 |
| Using scientific evidence | 509.4 | 9.50 | 458.8 | 33.53 | -50.56 | -9.93 | 0.152 |

NOTE: Bias is calculated as the estimate from the out-of-school test administration minus the estimate from the in-school test administration. The relative bias is calculated as the bias divided by the estimate from the in-school test administration.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

## B.5.3 Student-Level Scores: Original Sample

We performed the same set of comparisons, comparing during school and outside of school administration at the student level. That is, we compared the mean score for all students administered PISA in school with all students administered PISA out of school. Table B-31 presents results based on five score categories for students in original sample schools. The table presents estimates in each score category for the students in schools with in-school and out-of-school administration. There is no evidence of a significant difference between the mean scores of students

tested during school hours and those tested outside of school hours. Also, the absolute values of the relative bias for all score categories are less than 2 percent.

**Table B-31. Mean student score in original sample schools, for administrations during and outside of school hours, by score category: Fall 2006**

| Score category | In-school | | Out-of-school | | Bias | Relative bias | *p*-value for *t* test |
|---|---|---|---|---|---|---|---|
| | Estimate | Standard error | Estimate | Standard error | | | |
| Math | 479.2 | 2.75 | 470.4 | 3.77 | -8.75 | -1.83 | 0.058 |
| Science | 492.1 | 2.74 | 485.1 | 4.24 | -7.07 | -1.44 | 0.161 |
| Explaining phenomena scientifically | 489.0 | 2.72 | 482.5 | 4.25 | -6.52 | -1.33 | 0.183 |
| Identifying scientific issues | 492.0 | 2.58 | 491.5 | 3.89 | -0.51 | -0.10 | 0.912 |
| Using scientific evidence | 493.2 | 3.08 | 483.7 | 5.12 | -9.48 | -1.92 | 0.102 |

NOTE: Bias is calculated as the estimate from the out-of-school test administration minus the estimate from the in-school test administration. The relative bias is calculated as the bias divided by the estimate from the in-school test administration.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

### B.5.4 Student-Level Scores: Final Sample

Table B-32 presents results based on five score categories for students in final sample schools. The table presents estimates in each score category for the students in schools with in-school and out-of-school administration. There is no evidence of a significant difference between the mean scores of students tested during school hours and those tested outside of school hours. Also, the absolute values of the relative bias for all score categories are less than 2 percent.

**Table B-32. Mean student score in final sample schools, for administrations during and outside of school hours, by score category: Fall 2006**

| Score category | In-school | | Out-of-school | | Bias | Relative bias | *p*-value for *t* test |
|---|---|---|---|---|---|---|---|
| | Estimate | Standard error | Estimate | Standard error | | | |
| Math | 477.5 | 2.41 | 471.1 | 3.18 | -6.42 | -1.34 | 0.105 |
| Science | 491.8 | 2.39 | 485.8 | 3.55 | -5.98 | -1.22 | 0.172 |
| Explaining phenomena scientifically | 489.0 | 2.41 | 483.0 | 3.61 | -6.03 | -1.23 | 0.166 |
| Identifying scientific issues | 492.3 | 2.26 | 491.8 | 3.24 | -0.47 | -0.10 | 0.905 |
| Using scientific evidence | 492.2 | 2.70 | 484.6 | 4.27 | -7.53 | -1.53 | 0.132 |

NOTE: Bias is calculated as the estimate from the out-of-school test administration minus the estimate from the in-school test administration. The relative bias is calculated as the bias divided by the estimate from the in-school test administration.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

## B.6 Regression Analysis

### B.6.1 Student-level Scores, Final Sample

To supplement the above bivariate analyses we also carried out a multivariate regression analysis, with the student score (test scores and test subscales) as the dependent variable and selected school-level characteristics as the independent, explanatory variables. Table B-33 presents the results of this analysis, which was conducted using the final sample. This analysis allows study of the

relationship between test administration (during or outside of normal school hours) and scores on the assessment while controlling for other school characteristics as they may be related to scores. SUDAAN was used to test the statistical significance, and data were weighted using the final student weight.

**Table B-33. Results of regression analysis on student scores, by selected school characteristics: Fall 2006**

| School characteristics | Math | Science | Explaining phenomena scientifically | Identifying scientific issues | Using scientific evidence |
|---|---|---|---|---|---|
| Intercept | 547.42* | 577.78* | 576.89* | 576.08* | 583.27* |
| Public school | -53.67* | -56.50* | -52.67* | -54.93* | -65.85* |
| In-school administration | -4.84 | -5.17 | -5.22 | 0.80 | -6.48 |
| Minority | -18.47* | -29.85* | -31.16* | -23.43* | -28.15* |
| Central City | -22.65* | -23.38* | -24.43* | -19.51* | -28.90* |
| Urban fringe or large town | -2.16 | 2.31 | 0.30 | 0.57 | 4.45 |
| Northeast | -1.71 | -12.82* | -14.40* | -21.99* | -11.15 |
| South | -3.35 | -9.90* | -15.03* | -15.39* | -6.22 |
| Midwest | -27.62* | -34.65* | -35.68* | -35.92* | -38.43* |
| Total school enrollment | 0.04* | 0.04* | 0.04* | 0.03* | 0.03 |
| Age-eligible students | -0.09 | -0.09 | -0.13* | -0.09 | -0.07 |
| Students eligible for free or reduced-priced lunch | -0.00* | -0.00* | -0.00* | -0.00* | -0.00* |

* $p < .05$.
NOTE: Minority is a dichotomous variable. If a school has fewer than 15% of the enrollment as minorities (Black, Asian/Pacific Islander, American Indian/Alaska Native, Hispanic) then it is a non-minority school, and otherwise it is a minority school. See section B.9 for state listings within regions.
SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), 2006.

The coefficient for administration during school was not statistically significant in the model for any of the five scores. This adds confirmation to the results of bivariate tests indicating that administration during school does not have a significant relationship with final scores. (Note that the intercept is included in the table for the sake of completeness, but it is not substantively meaningful. By including it and showing it as significant, we show that the sample size is sufficient to accept the intercept in the model, and the model can be replicated by data users, if necessary.)

## B.7 Nonresponse Bias Analysis: Item

This section presents the results of the item nonresponse bias analysis. This analysis was conducted for the seven school questionnaire items with a response rate less than 85 percent and for the eight student questionnaire items with a response rate less than 85 percent. For each questionnaire item, respondents for that item were compared with nonrespondents for that item based on demographic characteristics known for everyone. These characteristics are from the CCD and PSS files, and continuous variables were made into categorical variables based on quartiles for the purpose of this analysis. For each category of each variable, bias was computed as the percentage of all item respondents who are in that category minus the percentage of all item nonrespondents who are in that category.

Table B-34 summarizes the school questionnaire item nonresponse bias analysis. The mean estimated bias across the seven questionnaire items and demographic characteristics ranges from 10.87 to 15.05, and the median estimated bias ranges from 8.15 to 12.50. The 39 bias estimates were tested for significance at the 0.05 level, and the percentage of the biases shown in table B-34 to be

statistically significant ranges from 7.69 to 17.95 with questionnaire item 6b[3] having the largest percentage of biases being significant. Five of the seven questionnaire items were significantly biased for public and private school types. There was no significant bias for any of the categories for the characteristics of total school enrollment, percent White student enrollment, and percent Other student enrollment.

**Table B-34. Estimated bias for school questionnaire items with a response rate less than 85 percent, by demographic characteristics: Fall 2006**

| Demographic characteristics | Estimated bias | | | | | | |
|---|---|---|---|---|---|---|---|
| | Q3c[1] | Q3d[2] | Q6a[3] | Q6b[4] | Q10a[5] | Q10b[6] | Q10c[7] |
| Census region | | | | | | | |
| Northeast | -28.93 | 0.18 | -25.43 | -11.80 | -5.14 | -0.25 | -1.01 |
| Midwest | 2.75 | -0.48 | 21.83 | 29.67* | 12.55 | 29.41* | -0.14 |
| South | 9.49 | -16.22 | -11.78 | -25.18 | -24.03 | -14.81 | -12.77 |
| West | 16.69* | 16.53 | 15.38 | 7.30 | 16.62* | -14.35 | 13.91 |
| Community type | | | | | | | |
| Central city | 5.87 | -8.61 | 11.04 | 8.87 | -25.83 | -42.95* | -13.83 |
| Urban fringe or large town | -25.53 | 0.63 | -2.89 | 12.36 | 11.52 | 12.94 | 11.82 |
| Rural or small town | 19.66 | 7.98 | -8.15 | -21.22 | 14.31 | 30.01* | 2.02 |
| School type | | | | | | | |
| Public school | -24.16* | -15.05 | -28.75* | -28.22* | -28.80* | 0.72 | -32.40* |
| Private school | 24.16* | 15.05 | 28.75* | 28.22* | 28.80* | -0.72 | 32.40* |
| Total school enrollment | | | | | | | |
| 1–629 | 21.43 | -8.64 | 17.12 | -0.12 | 6.28 | 3.56 | 3.85 |
| 630–1,152 | -1.18 | 4.51 | -5.81 | 10.94 | -3.21 | 1.76 | 1.90 |
| 1,153–1,850 | -15.45 | 0.81 | -6.80 | -11.48 | -1.85 | -2.90 | -3.22 |
| More than 1,850 | -4.80 | 3.31 | -4.51 | 0.66 | -1.22 | -2.42 | -2.53 |
| Percent Asian/Pacific Islander, non-Hispanic student enrollment | | | | | | | |
| 0–0.44 | 15.03 | -14.50 | 8.50 | -8.27 | -15.80 | 7.36 | -21.70 |
| 0.45–1.73 | -2.47 | -4.45 | 11.45 | 22.71* | 12.43 | 14.89 | 16.20 |
| 1.74–4.03 | -11.69 | 5.75 | -22.70 | -19.30 | -7.51 | -8.46 | -7.60 |
| 4.04–100 | -0.87 | 13.19 | 2.75 | 4.86 | 10.88 | -13.79 | 13.10 |
| Percent Black, non-Hispanic student enrollment | | | | | | | |
| 0–1.25 | -9.62 | -15.90 | 31.95 * | 31.00* | 12.48 | 23.89 | -4.45 |
| 1.26–4.91 | 12.99 | 25.89* | -27.33 | -27.93 | 19.61 | -3.92 | 22.04 |
| 4.92–20.47 | -0.88 | 12.01 | -1.78 | 16.83 | 8.38 | 8.42 | 11.05 |
| 20.48–100 | -2.50 | -22.00 | -2.83 | -19.89 | -40.48* | -28.40 | -28.64 |
| Percent White, non-Hispanic student enrollment | | | | | | | |
| 0–51.01 | 4.98 | -12.05 | -0.95 | -0.30 | -31.80 | -17.10 | -19.94 |
| 51.02–72.85 | -2.60 | 2.79 | -2.92 | 5.22 | 1.02 | -8.36 | -0.97 |
| 72.86–91.67 | 4.45 | 23.20 | 0.78 | -13.44 | 16.18 | -1.43 | 24.25 |
| 91.68–100 | -6.83 | -13.94 | 3.09 | 8.52 | 14.61 | 26.89 | -3.35 |

[3] Question 6b asks "About what percentage of students in your school repeated a grade at the high school level (grades 10–12), last academic year?"

**Table B-34. Estimated bias for school questionnaire items with a response rate less than 85 percent, by demographic characteristics: Fall 2006—Continued**

| Demographic characteristics | Estimated bias | | | | | | |
|---|---|---|---|---|---|---|---|
| | Q3c[1] | Q3d[2] | Q6a[3] | Q6b[4] | Q10a[5] | Q10b[6] | Q10c[7] |
| Percent American Indian/Alaska Native, non-Hispanic student enrollment | | | | | | | |
| 0–0.05 | 18.15 | -11.46 | 18.36 | 13.72 | -21.26 | -24.41 | -1.52 |
| 0.06–0.28 | -19.57 | 1.05 | -7.89 | 5.51 | -0.41 | -1.56 | 1.71 |
| 0.29–0.63 | -22.59 | 4.48 | -3.65 | -12.50 | -5.49 | -3.66 | -0.01 |
| 0.63–100 | 24.02 | 5.94 | -6.83 | -6.73 | 27.16* | 29.63* | -0.18 |
| Percent Hispanic student enrollment | | | | | | | |
| 0–1.50 | -26.44 | -43.54* | 26.82 | 27.50 | -11.03 | 11.90 | -21.06 |
| 1.51–4.65 | 13.65 | 22.66* | -24.62 | -30.11 | 8.62 | 15.13 | 18.20 |
| 4.66–15.64 | 12.86 | 17.78* | -2.53 | -3.48 | 7.26 | -17.09 | 10.00 |
| 15.65–100 | -0.07 | 3.10 | 0.33 | 6.09 | -4.86 | -9.95 | -7.14 |
| Percent other student enrollment | | | | | | | |
| Schools with other students enrolled | -0.68 | -3.61 | 5.41 | 6.02 | 1.28 | 2.99 | 2.15 |
| Schools without other students enrolled | 0.68 | 3.61 | -5.41 | -6.02 | -1.28 | -2.99 | -2.15 |
| Percent students eligible for free or reduced-price lunch | | | | | | | |
| 0–7.19 | 10.45 | 14.15 | 24.91* | 33.21* | 22.75* | -5.57 | 27.42* |
| 7.20–23.90 | -6.91 | -20.75 | -19.61 | -13.37 | 13.32 | 23.46 | -9.30 |
| 23.91–41.89 | -15.22 | 12.63 | 7.80 | 14.19* | -12.11 | -5.51 | -5.12 |
| 41.90–100 | 11.68 | -6.03 | -13.10 | -34.03 | -23.96 | -12.38 | -13.00 |
| Mean estimated bias[8] | 11.74 | 11.14 | 12.12 | 15.05 | 13.64 | 12.46 | 10.87 |
| Median estimated bias[8] | 11.68 | 11.46 | 8.15 | 12.50 | 12.43 | 9.95 | 9.30 |
| Percent significant bias | 7.69 | 10.26 | 10.26 | 17.95 | 15.38 | 10.26 | 7.69 |

* Bias is significant at the 0.05 level.

[1] Q3c: About what percentage of your total funding for a typical school year comes from benefactors, donations, bequests, sponsorships, parent fund raising?

[2] Q3d: About what percentage of your total funding for a typical school year comes from other sources?

[3] Q6a: About what percentage of students in your school repeated a grade at the middle or junior high school level (grades 7–9), last academic year?

[4] Q6b: About what percentage of students in your school repeated a grade at the high school level (grades 10–12), last academic year?

[5] Q10a: How many part-time teachers are on the staff of your school?

[6] Q10b: How many part-time teachers fully certified by the state in the main assignment field are on the staff of your school?

[7] Q10c How many part-time teachers with a bachelor's degree are on the staff of your school?

[8] Mean and median bias are based on the absolute values of the estimated bias.

NOTE: See section B.9 for state listings within regions. 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.

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

Table B-35 summarizes the student questionnaire item nonresponse bias analysis. The mean estimated bias across the eight questionnaire items and demographic characteristics ranges from 1.87 to 4.78, and the median estimated bias ranges from 1.30 to 4.22. The 39 bias estimates were tested for significance at the 0.05 level, and the percentage of the biases shown in table B-35 to be

statistically significant ranges from 64.10 to 79.49 with questionnaire item Q10a[4] having the largest percentage with significant bias. Seven of the eight items were significantly biased for the rural or small town community type, 0 to 1.25 percent Black student enrollment, 91.68 to 100 percent White student enrollment, and 0.29 to 0.63 percent American Indian/Alaska Native student enrollment. All eight items were significantly biased for the Census Midwest region, central city and urban fringe or large town community types, public and private school types, 1 to 629 student enrollment, 20.48 to 100 percent Black student enrollment, and 0 to 7.19 percent and 41.9 to 100 percent students eligible for free or reduced-price lunch.

[4] Question 10a asks "Does your father have a B.A. (4-year), master's, doctoral, or professional degree, such as medicine or law?"

**Table B-35. Estimated bias for student questionnaire items with a response rate less than 85 percent, by demographic characteristics: Fall 2006**

| Demographic characteristics | Estimated bias | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | MSE-CATEG[1] | Q7b[2] | Q7c[3] | FSE-CATEG[4] | Q10a[5] | Q10b[6] | Q10c[7] | Q30a[8] |
| Census region | | | | | | | | |
| Northeast | 1.44 | -5.35* | -5.13* | 2.43* | -2.74* | -3.77* | -4.33* | -2.26 |
| Midwest | 4.70* | 4.08* | 3.47* | 3.81* | 4.68* | 4.25* | 3.78* | 9.23* |
| South | -4.19* | -2.03* | -1.38 | -7.00* | -2.38 | -2.26* | -1.26 | -5.03* |
| West | -1.95 | 3.30* | 3.04* | 0.77 | 0.44 | 1.78* | 1.81* | -1.94 |
| Community type | | | | | | | | |
| Central city | -6.58* | -3.26* | -2.40* | -13.85* | -6.07* | -4.01* | -3.95* | -10.91* |
| Urban fringe or large town | 2.51* | -3.12* | -4.62* | 6.44* | 4.42* | -3.33* | -3.67* | 6.68* |
| Rural or small town | 4.07* | 6.39* | 7.02* | 7.41* | 1.65 | 7.34* | 7.62* | 4.23* |
| School type | | | | | | | | |
| Public school | -2.62* | 3.56* | 3.75* | -5.68* | -4.04* | 1.30* | 2.12* | -1.80* |
| Private school | 2.62* | -3.56* | -3.75* | 5.68* | 4.04* | -1.30* | -2.12* | 1.80* |
| Total school enrollment | | | | | | | | |
| 1–629 | 3.74* | 2.54* | 2.71* | 5.23* | 2.57* | 2.12* | 2.78* | 4.88* |
| 630–1,152 | -0.29 | -4.24* | -4.58* | 2.60 | 1.54 | -0.93 | -2.39* | -4.22 |
| 1,153–1,850 | 4.18* | 2.46* | 3.84* | -1.58 | 0.14 | 1.21 | 2.13* | 8.81* |
| More than 1,850 | -7.63* | -0.76 | -1.97 | -6.24* | -4.25* | -2.39* | -2.51* | -9.47* |
| Percent Asian/Pacific Islander non-Hispanic student enrollment | | | | | | | | |
| 0–0.44 | -3.45* | 3.84* | 3.34* | -3.61* | -1.31 | 2.46* | 3.71* | -3.30 |
| 0.45–1.73 | 5.81* | -2.67* | -2.16* | 4.07* | -0.02 | -0.67 | -1.12 | 3.62 |
| 1.74–4.03 | 2.84* | -0.09 | 0.05 | -0.26 | -1.29 | -0.40 | -0.32 | 4.41* |
| 4.04–100 | -5.20* | -1.08 | -1.23 | -0.21 | 2.62* | -1.39 | -2.27* | -4.72 |
| Percent Black, non-Hispanic student enrollment | | | | | | | | |
| 0–1.25 | 5.65* | 1.25 | 2.88* | 10.79* | 3.15* | 2.33* | 2.59* | 8.45* |
| 1.26–4.91 | 2.79* | 1.01 | 0.74 | 7.97* | 7.62* | 3.37* | 3.29* | 8.92* |
| 4.92–20.47 | -5.65* | 0.17 | -0.74 | -5.14* | -2.96* | 0.02 | -1.21 | -10.52* |
| 20.48–100 | -2.79* | -2.43* | -2.88* | -13.62* | -7.82* | -5.72* | -4.67* | -6.85* |
| Percent White, non-Hispanic student enrollment | | | | | | | | |
| 0–51.01 | -10.62* | 0.18 | 1.14 | -16.39* | -8.06* | -2.52* | -1.96* | -9.40* |
| 51.02–72.85 | -1.44 | 1.73* | 0.46 | -7.44* | -5.90* | -0.94 | -1.39 | -8.57* |
| 72.86–91.67 | 3.46* | -3.28* | -3.86* | 10.59* | 6.11* | -0.79 | -1.17 | 6.13* |
| 91.68–100 | 8.59* | 1.36 | 2.26* | 13.25* | 7.85* | 4.25* | 4.52* | 11.84* |

See notes at end of table.

**Table B-35. Estimated bias for student questionnaire items with a response rate less than 85 percent, by demographic characteristics: Fall 2006—Continued**

| Demographic characteristics | Estimated bias | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | MSE-CATEG[1] | Q7b[2] | Q7c[3] | FSE-CATEG[4] | Q10a[5] | Q10b[6] | Q10c[7] | Q30a[8] |
| Percent American Indian/Alaska Native non-Hispanic student enrollment | | | | | | | | |
| 0–0.05 | 1.11 | 0.79 | 1.30 | 3.71* | 5.33* | 0.35 | 1.34 | 4.39* |
| 0.06–0.28 | 0.55 | -4.54* | -4.66* | -2.12 | -3.34* | -2.99* | -3.46* | 2.29 |
| 0.29–0.63 | -0.96 | -2.37* | -2.90* | -5.07* | -6.37* | -3.53* | -3.47* | -4.91* |
| 0.63–100 | -0.69 | 6.11* | 6.26* | 3.48* | 4.38* | 6.17* | 5.60* | -1.77 |
| Percent Hispanic student enrollment | | | | | | | | |
| 0–1.50 | 0.54 | -0.13 | 1.96 | 3.13 | 2.98* | 1.06 | 3.03* | -0.05 |
| 1.51–4.65 | 6.82* | 0.27 | -0.54 | 7.38* | 6.47* | 0.88 | 0.02 | 12.92* |
| 4.66–15.64 | 2.08* | 0.03 | 0.30 | -0.04 | -2.74* | -2.43* | -1.27 | 2.18 |
| 15.65–100 | -9.44* | -0.18 | -1.72 | -10.47* | -6.71* | 0.49 | -1.79* | -15.04* |
| Percent other student enrollment | | | | | | | | |
| Schools with other students enrolled | 0.17 | -2.57* | -3.09* | -0.34 | -2.31* | -0.91* | -1.24* | -2.00 |
| Schools without other students enrolled | -0.17 | 2.57* | 3.09* | 0.34 | 2.31* | 0.91* | 1.24* | 2.00 |
| Percent students eligible for free or reduced-price lunch | | | | | | | | |
| 0–7.19 | 5.06* | -9.85* | -9.53* | 8.90* | 3.47* | -5.62* | -7.53* | 4.33* |
| 7.20–23.90 | 3.85* | 0.80 | -0.32 | 9.08* | 5.55* | 2.48* | 0.28 | 3.97 |
| 23.91–41.89 | -0.44 | 4.88* | 5.36* | -5.16* | -3.33* | 1.55 | 3.45* | -1.08 |
| 41.90–100 | -8.47* | 4.17* | 4.48* | -12.82* | -5.70* | 1.59* | 3.80* | -7.22* |
| Mean estimated bias[9] | 2.96 | 2.10 | 2.35 | 4.78 | 3.16 | 1.87 | 2.17 | 4.53 |
| Median estimated bias[9] | 2.62 | 1.73 | 2.16 | 3.81 | 2.96 | 1.30 | 1.96 | 4.22 |
| Percent significant bias | 69.23 | 64.10 | 64.10 | 74.36 | 79.49 | 66.67 | 74.36 | 64.10 |

* Bias is significant at the 0.05 level.

[1] MSECATEG: Mother's white collar/blue collar classification—derived.

[2] Q7b: Does your mother have an associate's degree (2-year degree)?

[3] Q7c: Does your mother have vocational or technical qualifications?

[4] FSECATEG: Father's white collar/blue collar classification—derived.

[5] Q10a: Does your father have a B.A. (4-year), master's, doctoral, or professional degree, such as medicine or law?

[6] Q10b: Does your father have an associate's degree (2-year degree)?

[7] Q10c: Does your father have vocational or technical qualifications?

[8] Q30a: What kind of job do you expect to have when you are about 30 years old?

[9] Mean and median bias are based on the absolute values of the estimated bias.

NOTE: See section B.9 for state listings within regions. 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.

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

## B.8 Conclusions

This report presents nonresponse bias analysis results for both the original U.S. PISA 2006 sample and the final sample using a variety of statistical tests including chi-square, *t* test, and logistic regression. For the original sample there is only evidence of some significant nonresponse bias before weight adjustment with respect to the Other race/ethnicity composition of the responding schools. After weight adjustment, the significant bias remains limited to the Other students with a

bias of -0.8 percent indicating that the mean percentage enrollment of Other students is higher in responding schools. In addition, the relative bias before weight adjustment indicates the potential for bias in school control, community type, Census region, enrollment, race/ethnicity, and free or reduced-price lunch status. After weight adjustment, the relative bias is about 14 percent for the central city and 12 percent for Blacks with the estimates higher for respondents, and the relative bias is about 25 percent for the West Census region and 38 percent for American Indians/Alaska Natives with the estimates lower for respondents. Also, the logistic regression, which compared respondents and nonrespondents before weight adjustments, indicates potential bias for the percentage of Black, American Indian, and Asian students, which had higher estimates for nonrespondents.

For the final sample, that is with respondents redefined as including substitute schools, the analysis before weight adjustment suggests the presence of significant nonresponse bias only with respect to the Other race/ethnicity composition of the responding schools indicating that the mean percentage enrollment of Other students is higher in responding schools. However, after weight adjustment, no significant bias remains for the Other race/ethnicity, but the Hispanic category becomes significant indicating that the mean percentage enrollment of Hispanic students is higher in responding schools. In addition, the relative bias before weight adjustment indicates the potential for bias in school control, community type, Census region, enrollment, race/ethnicity, and free or reduced-price lunch status. After weight adjustment, the relative bias is about 12 percent for the central city and 13 percent for Other race/ethnicity with the estimates higher for respondents, and the relative bias is about 31 percent for the West Census region, and 49 percent for American Indians/Alaska Natives with the estimates lower for respondents. Also, the logistic regression, which compared respondents and nonrespondents before weight adjustments, indicates potential bias for publicly controlled schools, urban fringe or large town, the percentage of White and American Indian students, which had higher estimates for nonrespondents, and the South Census region, which had higher estimates for respondents.

Overall, the data suggest that while there is minimal significant nonresponse bias and the biases are generally reduced after weight adjustment, large relative biases exist before and after weight adjustments. The bias in the released data is generally not large, but there is potential nonresponse bias in several variables. Therefore, data users should use caution when analyzing the data, especially when the analysis involves variables identified in this report as being subject to nonresponse bias.

No significant differences were found between the schools that administered the test during out-of-school hours and the schools that opted for traditional in-school testing. Tests for differences in a variety of school characteristics demonstrated no significant results. Tests for differences in student test scores were implemented at both the school and student levels, and no significant difference was found between the two groups of schools. Finally, a regression analysis of scores as a function of selected school characteristics resulted in no evidence of a significant effect of the type of administration on the final test scores.

The item nonresponse bias analysis found that mean and median estimated biases are less than 16 percentage points for school items and less than 5 percentage points for student items. However, for each variable there were significant differences in response rates for at least one demographic group tested.

## B.9 Technical Notes

### B.9.1 Description of Variables

Frame characteristics for public schools were taken from the 2003–04 Common Core of Data (CCD) and, for private schools, from a preliminary version of the 2003–04 Private School Universe Survey (PSS). The CCD and PSS were used to categorize schools as public or private.

### B.9.2 Race/Ethnicity

The percentage of students of various race/ethnicity categories was taken from the sampling frame, based on variables collected in the CCD and PSS. The categories used were percent of students who were Asian/Pacific Islander, Black non-Hispanic, Hispanic, American Indian/ Alaska Native, White non-Hispanic, and "Other."

### B.9.3 Community Type

Community type is based on the school's location relative to populous areas (the school's location is based on its address).

- Central city consists of *large city* (a principal city of a metropolitan core-based statistical area [CBSA], with the city having a population greater than or equal to 250,000) and *midsize city* (a principal city of a metropolitan CBSA, with the city having a population less than 250,000).
- Urban fringe/large town consists of *urban fringe of a large city* (any incorporated place, Census-designated place, or nonplace territory within a metropolitan CBSA of a large city and defined as urban by the Census Bureau), *urban fringe of a midsize city* (any incorporated place, Census-designated place, or nonplace territory within a CBSA of a midsize city and defined as urban by the Census Bureau), and *large town* (any incorporated place or Census-designated place with a population greater than or equal to 25,000 and located outside a metropolitan CBSA or inside a micropolitan CBSA).
- Rural/small town consists of *small town* (any incorporated place or Census-designated place with a population less than 25,000 and greater than or equal to 2,500 and located outside a CBSA or CSA (consolidated statistical area); *rural, outside CBSA* (any incorporated place, Census-designated place, or nonplace territory not within a CBSA or CSA and defined as rural by the Census Bureau); and *rural, inside CBSA* (any incorporated place, Census-designated place, or nonplace territory within a metropolitan CBSA and defined as rural by the Census Bureau).

### B.9.4 Census Region

Census region consists of the following divisions:

- Northeast—Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont;
- Midwest—Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin;

- West—Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming; and
- South—Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia.

## B.9.5 Poverty Level in Public Schools (Percentage of Students Eligible for Free or Reduced-Price Lunch)

For sampling (see section B.9.6.1), schools were grouped into high poverty—schools in which 50 percent or more of students were eligible to receive free or reduced-price lunch through the National School Lunch Program—and low poverty—schools in which less than 50 percent of students were eligible. All private schools were classified as low poverty. This was implemented to prevent excessive oversampling of schools that were both private and high poverty. Also, information on poverty status for private schools was not available on the frame.

Information on poverty level used in this report was obtained from principals' responses to the school questionnaire. The questionnaire asked what percentage of students at the school was eligible to receive free or reduced-price lunch through the National School Lunch Program. Analyses included in this report used the original, continuous version of this variable, namely, the proportion of students eligible to receive free or reduced-price lunch, or a categorical version of the continuous variable split into quartiles.

## B.9.6 Statistical Procedures

### B.9.6.1 Sampling

The sampling frame for the U.S. school sample was constructed using data from the 2002–03 Common Core of Data (CCD) and preliminary data from the 2003–04 Private School Universe Survey (PSS). Before the selection process, schools were sorted into two explicit groups: large schools and small schools. The schools were then sorted by implicit strata: grade span of school, percentage of racial/ethnic minority students, control of school (public/private), percentage of students eligible for free or reduced-priced lunch, and locale. Schools were selected on the basis of the number of 15-year-old students in the school, so that schools with more students in 9th and 10th grades had a higher probability of selection than schools with fewer grades containing 15-year-olds. The final sample included 236 schools.

### B.9.6.2 Weighting

Records from the sample schools and students were assigned sampling (design) weights to adjust for over- or underrepresentation from a particular group. The use of design weights is necessary for the computation of statistically sound, nationally representative estimators. The weight assigned to a school's (or student's) data is the inverse of the probability that the school (or student) would be selected for the sample. When data are weighted, each sample unit contributes to the results in proportion to the total number of schools or students represented by that unit. A school-level participation (nonresponse) adjustment was then made in the school weight to compensate for any sampled schools that did not participate and were not replaced. That adjustment was calculated independently for each explicit stratum. That adjustment was calculated independently for each explicit stratum described in A.6.1. The weight after nonresponse adjustment is the final weight.

### B.9.6.3 Sampling Errors

Sampling errors occur when the discrepancy between a population characteristic and the sample estimate arises because not all members of the reference 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. This particular sample of 15-year-old students from the 2005–06 school year was just one of many possible samples that could have been selected. Therefore, estimates produced from the PISA sample may differ from estimates that would have been produced had another student sample been drawn. This type of variability is called sampling error because it arises from using a sample of the population, rather than all of its members.

The standard error is a measure of the variability caused by sampling when estimating a statistic, and is often included in reports containing estimates from survey data. The approach used for calculating sampling variances in PISA was the Taylor Series expansion. In this report we do not show estimates of standard errors for each estimate. Rather, the effects of sampling error are reflected in the test statistics that are presented for each analysis. These are described below.

### B.9.6.4 Tests of Significance

Comparisons made in the text of this report have been tested for statistical significance. For example, when comparing results obtained from the responding sample for a given grade with those obtained from the nonresponding sample units, tests of statistical significance were used to establish whether the observed differences are statistically significant. The estimation of the standard errors that are required to undertake the tests of significance requires incorporation of the complex sample design.

Two kinds of statistical tests are included in the report: *t* tests and chi-square tests. In addition, logistic regression analyses were conducted.

### B.9.6.5 *T* Tests

*T* tests were used to test for the hypothesis that no difference exists between the means of continuous variables for two groups (namely, the responding sample and the nonresponding sample). Suppose that $\bar{x}_A$ and $\bar{x}_B$ are the means for two groups that are being compared, and $se(\bar{x}_A - \bar{x}_B)$ is the standard error of the difference between the means that accounts for the complex survey design. Then the *t* test is defined as

$$t = \frac{|\bar{x}_A - \bar{x}_B|}{se(\bar{x}_A - \bar{x}_B)}$$

This statistic is then compared with the critical values of the appropriate student *t* distribution to determine whether the difference is statistically significant. The appropriate number of degrees of freedom for the distribution is given by the number of primary sampling units in the design (in this case, the number of schools) minus the number of sampling strata.

### B.9.6.6 Chi-Square Tests

The Wald F statistic based on the Wald chi-square test was used to test whether two distributions of a given categorical variable are different. Using SUDAAN, this testing was conducted in a way that reflects the impact of the complex sample design on sampling variance.

### B.9.6.7 Logistic Regression Models

A linear model for investigating the relationship between binary (dichotomous) outcomes and a set of explanatory variables is referred to as a *logistic regression model*. The data are assumed to follow a binomial distribution, with probabilities that depend on the independent variables. Let $p_i$ denote the probability that the $i$th sampled school will respond. Under the logistic regression model, the log odds of response propensity (expressed in terms of the logarithm of $p_i/(1-p_i)$) are assumed to have the following linear form:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$$

where $X_{1i}$, $X_{2i}$ …, $X_{pi}$ are $p$ auxiliary variables associated with the $i$th sampled beneficiary, and $\beta_0$, $\beta_1$, ..., $\beta_p$ are coefficients to be estimated. Asymptotic assumptions are used to develop statistical tests to determine which, if any, of the coefficients are significantly different from zero. In the analyses in this report, the standard procedures for carrying out logistic regression analyses have been modified both to incorporate the sampling weights in the estimation of the coefficients and to reflect the effect of the complex sample design on the variance-covariance matrix of the coefficients.

# Appendix C.
# Endorsing Organizations

## Endorsing Organizations

American Association of School Administrators

American Federation of Teachers

Council for American Private Education

Council of Chief State School Officers

International Reading Association

National Association of Elementary School Principals

National Association of Independent Schools

National Association of Secondary School Principals

National Catholic Educational Association

National Christian School Association

National Education Association

National Parent Teacher Association

National School Board Association

National Science Teachers Association

# Appendix D.
# Recruitment Materials

## D.1 State Notification Letter

U.S. DEPARTMENT OF EDUCATION
INSTITUTE OF EDUCATION SCIENCES
NATIONAL CENTER FOR EDUCATION STATISTICS

[DATE]

[NAME]
[STATE DEPARTMENT OF EDUCATION]
[ADDRESS]

[CITY, STATE ZIP]

Dear [CHIEF STATE SCHOOL OFFICER]:

I am writing to inform your state education agency about an upcoming international study: the Program for International Student Assessment (PISA). PISA is an international study designed to provide internationally comparable information about student performance in several academic subjects. Students in about 60 countries, including the United States, are participating in PISA. Benchmarking the performance of the United States in relation to other countries is an important measure of our nation's progress in educating all of our children. PISA is sponsored in the United States by the National Center for Education Statistics, U.S. Department of Education, and conducted by RTI International (RTI), a non-profit research organization in North Carolina.

PISA provides a unique opportunity to evaluate the knowledge and skills of 15-year-old students throughout the world. By comparing our students' performance with that of students in other nations, we can see where we are successful and where we still face challenges in educating our youth. Data will be collected to assess proficiency in science, mathematics, and reading, with particular focus on science proficiency.

We ask your agency to support the participation of districts and schools in your state in the PISA study. We will be conducting the main study for PISA in the fall of 2006. One or more public schools in your state have been randomly sampled to participate in the PISA study. More information about the study can be found in the enclosed materials.

Within the next few days, we plan to contact sampled school districts and schools to discuss conducting data collection in the sampled schools for PISA. In the meantime, if you have questions about the study, please call Dr. Patricia Green at RTI at (877) 225-0771. For more information about PISA, you may contact Dr. Elois Scott at NCES at (202) 502-7489 or visit the PISA website: http://nces.ed.gov/surveys/pisa/.

Thank you for your support of PISA.

Sincerely,

Val Plisko

Val Plisko
Associate Commissioner, Early Childhood, International & Crosscutting Studies
National Center for Education Statistics

cc: [STATE ASSESSMENT COORDINATOR]

WASHINGTON, D.C. 20006–

## D.2 District Notification Letter

U.S. DEPARTMENT OF EDUCATION
INSTITUTE OF EDUCATION SCIENCES
NATIONAL CENTER FOR EDUCATION STATISTICS

[DATE]

[SUPERINTENDENT NAME]
[DISTRICT NAME]
[ADDRESS]
[CITY, STATE ZIP]

Dear [SUPERINTENDENT]:

I am writing to inform your school district about an upcoming international study: the Program for International Student Assessment (PISA). PISA is an international study designed to provide internationally comparable information about student performance in several academic subjects. PISA focuses on the science literacy, mathematics literacy, and reading literacy of 15-year-old students. Students in about 60 countries, including the United States, are participating in PISA. Benchmarking the performance of the United States in relation to other countries is an important measure of our nation's progress in educating all of our children. PISA is sponsored in the United States by the National Center for Education Statistics (NCES), U.S. Department of Education, and conducted by RTI International (RTI), a non-profit research organization in North Carolina.

PISA provides a unique opportunity to evaluate the knowledge and skills of 15-year-old students throughout the world. By comparing our students' performance with that of students in other nations, we can see where we are successful and where we still face challenges in educating our youth. Data will be collected to assess proficiency in science, mathematics, and reading, with particular focus on science literacy.

We ask you to support the participation of schools in your district in the PISA study. Information about districts, students, and schools sampled for PISA is protected by the Education Sciences Reform Act of 2002 (PL 107-279). We will disclose the names of schools in each district only to the governing district for each school, and we ask that each district also maintain the confidentiality of the sampled schools in PISA. In the next few days, we will contact the following schools in your district which have been selected for PISA: [LIST ORIGINAL SCHOOLS HERE]. We may also contact: [LIST SUBSTITUTE SCHOOLS HERE]

We will be conducting the main study for PISA in the fall of 2006. Study reports will not identify participating districts, schools, students, parents or individual staff. More information about the study and next steps can be found in the enclosed materials.

If you have any questions about PISA or your district's participation in the study, please call Dr. Patricia J. Green at RTI at (877) 225-0771. For more information about PISA, you may contact Dr. Elois Scott at NCES at (202) 502-7489 or visit the PISA website: http://nces.ed.gov/surveys/pisa/.

Thank you for your support of PISA.

Sincerely,

Val Plisko

Val Plisko
Associate Commissioner, Early Childhood, International & Crosscutting Studies
National Center for Education Statistics

WASHINGTON, D.C. 20006–

## D.3 School Recruitment Letter

U.S. DEPARTMENT OF EDUCATION
INSTITUTE OF EDUCATION SCIENCES
NATIONAL CENTER FOR EDUCATION STATISTICS

[DATE]

[PRINCIPAL NAME]
[SCHOOL NAME]
[ADDRESS 1]
[CITY, STATE ZIP]

Dear [PRINCIPAL]:

I am writing to request your school's support for the Program for International Student Assessment (PISA). PISA is an international study designed to provide internationally comparable information about student performance in several academic subjects. PISA focuses on the science literacy, mathematics literacy, and reading literacy of 15-year-old students. Students in about 60 countries, including the United States, are participating in PISA. Benchmarking the performance of the United States in relation to other countries is an important measure of our nation's progress in educating all of our children. PISA is sponsored in the United States by the National Center for Education Statistics (NCES), U.S. Department of Education, and conducted by RTI International (RTI), a non-profit research organization in North Carolina.

PISA provides a unique opportunity to evaluate the knowledge and skills of 15-year-old students throughout the world. By comparing our students' performance with that of students in other nations, we can see where we are successful and where we still face challenges in educating our youth. Data will be collected to assess proficiency in science, mathematics, and reading, with particular focus on science proficiency.

We will be conducting the PISA main study in the fall of 2006. Study reports will not identify participating districts, schools, students, parents or individual staff. International comparisons are an extremely important part of monitoring educational performance in the United States. Each school selected for the PISA study contributes to providing a valid representation of the performance of U.S. students and is critical to the success of the study. More information about the study and next steps can be found in the enclosed materials.

Within a few days, a representative of RTI will call you to discuss any questions that you may have. In the meantime, if you have any questions please call Dr. Patricia J. Green at RTI at (877) 225-0771. You may also contact Dr. Elois Scott at NCES at (202) 502-7489 or visit the PISA website: http://nces.ed.gov/surveys/pisa/.

Thank you for your support of PISA.

Sincerely,

Val Plisko

Val Plisko
Associate Commissioner, Early Childhood, International & Crosscutting Studies
National Center for Education Statistics

WASHINGTON, D.C. 20006–

# Program for International Student Assessment (PISA) Fact Sheet

- Are students well prepared to meet the challenges of the future?
- Are they able to analyze, reason, and communicate their ideas effectively?
- Do they have the capacity to continue learning throughout life?
- How do U.S. students compare to their peers in other countries?

These are just some of the questions that the **Program for International Student Assessment (PISA)** hopes to answer. This fact sheet gives some important information about the study.

### What is PISA?

PISA is an international assessment that is designed by participating countries and administered to 15-year-olds in schools around the world. PISA reports on performance in reading literacy, mathematics literacy, and science literacy every 3 years, with emphasis on one subject in each data collection cycle. PISA focused on mathematics literacy in 2003 and reading literacy in 2000. In 2006, science literacy will be the primary focus.

### Why is participation important?

PISA 2006 takes place at a time when interest is increasing, both worldwide and in the United States., in how well schools are preparing students to meet the challenges of the future. In light of the growing concerns related to international economic competitiveness, the changing face of our workplace, and the expanding international marketplace we trade in, knowing how our students and adults compare with their peers around the world has become a more prominent issue than ever before. PISA is designed to further our understanding of how well other nations are advancing the educational achievement of their populations.

### When will the test administration be conducted?

The test administration period is between September 25 and November 22, 2006. We will work with each school to schedule a date convenient for the school. We will send a trained Test Administrator to the school to administer the student assessment.

### Do schools, school staff, and students have to participate?

Although participation in PISA is voluntary, it is important that every selected school and student participate to ensure the completeness and accuracy of results. Development of national results, and inclusion in the international comparisons, depends on a high response rate.

### How many countries participate?

In 2000, the U.S. was one of 32 countries to participate in the first PISA assessment. Forty-one countries participated in PISA 2003. Approximately 60 countries are currently involved in the design of PISA 2006, including: Argentina, Australia, Austria, Azerbaijan, Belgium, Brazil, Bulgaria, Canada, Chile, Chinese Taipei (Taiwan), Colombia, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hong Kong Special Administrative Region of China, Hungary, Iceland, Indonesia, Ireland, Israel, Italy, Japan, Jordan, Kazakhstan, Kyrgyzstan, Latvia, Liechtenstein, Lithuania, Luxembourg, Macao Special Administrative Region of China, Mexico, Montenegro, Netherlands, New Zealand, Norway, Peru, Poland, Portugal, Qatar, Republic of Korea, Romania, Russian Federation, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Thailand, Tunisia, Turkey, United Kingdom, United States, and Uruguay.

### Are there released test items we can review?

Yes, you can see all of the released test items and scoring guide for the PISA 2003 assessment by going to http://www.pisa.oecd.org/document/38/0,2340,en_32252351_32236173_34993126_1_1_1_1,00.html.

### How many schools, students, and staff participate?

The main study of PISA 2006 will consist of approximately 150 schools. We will select an average of 35 15-year-old students in each school to participate in the study. One school administrator will be asked to complete the school questionnaire.

### Will the responses of participants be kept confidential?

By law, all responses that relate to or describe identifiable characteristics of individuals may be used *only* for statistical purposes and may not be disclosed, or used, in identifiable form for any other purpose, unless otherwise compelled by law.

### How long will it take to complete the questionnaires and tests?

The PISA student survey and test will take approximately 160 minutes to complete. The school questionnaire will take about 30 minutes.

We recognize that instructional time is valuable and we want to add to the students' learning experience. Each participating student will receive $15 as a token of our appreciation. In addition, we offer an honorarium of $100, with an opportunity to earn up to an additional $50 for high student participation, to the designated coordinator at each school in appreciation for his/her efforts toward making the study successful.

### What are the benefits of participating?

The United States as a whole benefits from the contribution of each school and student toward the national picture of what 15-year-olds know about science, mathematics, and reading, and how they compare with 15-year-olds worldwide. Each participating school contributes to this larger picture and helps ensure that the results for the United States are truly representative of performance and variation across all types of communities and all types of students.

### Who endorses PISA?

PISA 2006 is endorsed by the following organizations: American Association of School Administrators, American Federation of Teachers, Council for American Private Education, Council of Chief State School Officers, International Reading Association, National Association of Elementary School Principals, National Association of Independent Schools, National Association of Secondary School Principals, National Catholic Educational Association, National Christian School Association, National Education Association, National Parent Teacher Association, National School Board Association, and the National Science Teachers Association.

### Who sponsors PISA?

PISA is sponsored internationally by the Organization for Economic Cooperation and Development (OECD), and sponsored in the United States by the National Center for Education Statistics (NCES), U.S. Department of Education. NCES has contracted RTI International, a non-profit research organization in North Carolina, to carry out the data collection in the United States.

### How may I obtain more information?

For more information, you may contact the project director, Dr. Patricia Green at RTI toll-free at (877) 225-0771 between 9AM and 5PM Eastern, or Dr. Elois Scott at NCES at (202) 502-7489. You may also visit the PISA website: http://nces.ed.gov/surveys/pisa.

## D.4 Implicit Consent Letter and Form

U.S. DEPARTMENT OF EDUCATION
INSTITUTE OF EDUCATION SCIENCES
NATIONAL CENTER FOR EDUCATION STATISTICS

Dear Madam or Sir:

We are pleased to inform you that your teenager has been selected to represent the United States in an important international study called the Program for International Student Assessment, or PISA. PISA is sponsored internationally by the Organization for Economic Cooperation and Development (OECD). The United States component is sponsored by the National Center for Education Statistics, U. S. Department of Education, and conducted by RTI International (RTI), a non-profit research organization in North Carolina.

The purpose of the study is to measure student learning in science, mathematics, and reading, and to compare progress of student learning in the United States to students in over 50 other countries. Benchmarking the performance of the United States in relation to other countries is an important measure of our nation's progress in educating all of our children. In a few weeks, your teenager will be asked to spend about three hours to complete the questionnaire and test along with approximately 35-45 other students in his/her school. Your teenager is one of only about 4,500 students from about 150 schools participating in PISA during the fall term of 2006.

This is an important opportunity for students to realize the value of science, mathematics and reading. The world wants to know how well 15-year-old students perform in these areas, and results of this study may help all educators and students in the future. Your teenager was selected to represent many others and cannot be replaced. We encourage your teenager to participate. We need your help to make PISA successful in the United States. We will provide $15 to each participating student as a token of our appreciation, and each student will receive a certificate commemorating his/her representation of the United States in PISA.

Information about districts, students, and schools sampled for PISA is protected by the Education Sciences Reform Act of 2002 (PL 107-279). By law, researchers may use the data for statistical purposes only. Data will be combined to produce statistical reports for Congress and others. No individual data (for example, names or addresses) will be reported. Participation is voluntary and there is no penalty if you or your teenager decides not to participate. Your teenager may choose not to answer any question. There are no risks to your teenager from taking part in the study.

If you are willing to allow your teenager to participate, you do not need to return this form. If for any reason you object to your teenager's participation, please fill out the enclosed form and return it to his/her school as soon as possible.

The enclosed brochure gives more information about the study. If you have any questions about PISA or your teenager's participation in the study, please call Cathy Forstner at RTI, toll-free, at 1-877-225-0771 between 9 AM and 5 PM Eastern time, Monday through Friday, or email us at PISA@rti.org. You may also visit the study website: http://nces.ed.gov/surveys/pisa/. If you have questions about your teenager's rights as a study participant, you may call RTI's Office for Research Protection, toll-free, at 1-866-214-2043. Both Ms. Forstner and staff from the Office for Research Protection can be reached at: RTI International, P.O. Box 12194, Research Triangle Park, NC 27709.

We thank you in advance for your cooperation in this important international study.

Sincerely,

Val Plisko

Val Plisko
Associate Commissioner, Early Childhood, International & Crosscutting Studies
National Center for Education Statistics

WASHINGTON, D.C. 20006–

## Implicit Consent Form

**PROGRAM FOR INTERNATIONAL STUDENT ASSESSMENT (PISA) PERMISSION FORM**

**IF YOU GRANT YOUR PERMISSION FOR YOUR TEENAGER TO PARTICIPATE IN THE STUDY, YOU DO NOT NEED TO RETURN THIS FORM.**

**IF YOU DO NOT CONSENT TO YOUR TEENAGER'S PARTICIPATION IN PISA, PLEASE RETURN THIS FORM TO YOUR TEENAGER'S SCHOOL AS SOON AS POSSIBLE.**

**I DO NOT GRANT PERMISSION** for my teenager, _____, to participate in the Program for International Student Assessment.

_____

(Signature of parent or guardian)

Date of signature: _____

(_____)_____

Area code Telephone number

PLEASE PRINT:

**Student name:** _____

School Name: _____

FOR OFFICE USE ONLY:

Student ID: _____

# Appendix E.
# Scoring Reliability Tables

**Table E-1. Science read-behind percent agreement, by scorer ID and science cluster: 2006**

| Science Cluster 1 | Scorer ID 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 26 | 30 | 29 | 30 | 27 | 30 | 22 | 24 | 24 | 24 | 29 | 30 | 28 | 30 | 48 | 48 | 15 | 15 |
| GMC 1 | 5 | 5 | 5 | 5 | 4 | 5 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 8 | 8 | 2 | 2 |
| ET 1 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 8 | 8 | 2 | 2 |
| ET 3 | 4 | 5 | 5 | 5 | 5 | 5 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 8 | 8 | 2 | 2 |
| TGC 1 | 4 | 5 | 5 | 5 | 4 | 5 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 8 | 8 | 3 | 3 |
| AR 2 | 4 | 5 | 4 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 8 | 8 | 3 | 3 |
| AR 5 | 4 | 5 | 5 | 5 | 4 | 5 | 3 | 4 | 4 | 4 | 4 | 5 | 3 | 5 | 8 | 8 | 3 | 3 |
| Percent agreement by evaluator | | 86.7 | | 96.7 | | 90.0 | | 91.7 | | 100.0 | | 96.7 | | 93.3 | | 100.0 | | 100.0 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 14 | 15 | 15 | 15 | 14 | 15 | 29 | 34 | 15 | 17 | 14 | 14 | 14 | 14 | 363 | 385 | 94.3 |
| GMC 1 | 2 | 2 | 2 | 2 | 2 | 2 | 5 | 6 | 1 | 2 | 2 | 2 | 2 | 2 | 58 | 61 | 95.1 |
| ET 1 | 2 | 2 | 2 | 2 | 2 | 2 | 5 | 5 | 3 | 3 | 2 | 2 | 2 | 2 | 61 | 61 | 100.0 |
| ET 3 | 1 | 2 | 2 | 2 | 2 | 2 | 5 | 5 | 3 | 3 | 2 | 2 | 2 | 2 | 58 | 61 | 95.1 |
| TGC 1 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 6 | 3 | 3 | 2 | 2 | 2 | 2 | 63 | 66 | 95.5 |
| AR 2 | 3 | 3 | 3 | 3 | 2 | 3 | 4 | 6 | 3 | 3 | 3 | 3 | 3 | 3 | 63 | 68 | 92.6 |
| AR 5 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 6 | 2 | 3 | 3 | 3 | 3 | 3 | 60 | 68 | 88.2 |
| Percent agreement by evaluator | | 93.3 | | 100.0 | | 93.3 | | 85.3 | | 88.2 | | 100.0 | | 100.0 | | 94.3 | † |

See notes at end of table.

**Table E--1. Science read-behind percent agreement, by scorer ID and science cluster: 2006—Continued**

| Science Cluster 2 | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 46 | 48 | 31 | 32 | 30 | 32 | 30 | 32 | 16 | 16 | 16 | 16 | 23 | 24 | 31 | 32 | 20 | 20 |
| GH 3 | 12 | 12 | 8 | 8 | 7 | 8 | 8 | 8 | 4 | 4 | 4 | 4 | 5 | 6 | 8 | 8 | 5 | 5 |
| GH 4 | 10 | 12 | 8 | 8 | 7 | 8 | 8 | 8 | 4 | 4 | 4 | 4 | 6 | 6 | 7 | 8 | 5 | 5 |
| GH 5 | 12 | 12 | 7 | 8 | 8 | 8 | 6 | 8 | 4 | 4 | 4 | 4 | 6 | 6 | 8 | 8 | 5 | 5 |
| RAD 3 | 12 | 12 | 8 | 8 | 8 | 8 | 8 | 8 | 4 | 4 | 4 | 4 | 6 | 6 | 8 | 8 | 5 | 5 |
| Percent agreement by evaluator | | 95.8 | | 96.9 | | 93.8 | | 93.8 | | 100.0 | | 100.0 | | 95.8 | | 96.9 | | 100.0 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 19 | 20 | 22 | 24 | 16 | 16 | 33 | 36 | 15 | 16 | 16 | 16 | 15 | 16 | 379 | 396 | 95.7 |
| GH 3 | 5 | 5 | 6 | 6 | 4 | 4 | 9 | 9 | 4 | 4 | 4 | 4 | 3 | 4 | 96 | 99 | 97.0 |
| GH 4 | 5 | 5 | 4 | 6 | 4 | 4 | 8 | 9 | 4 | 4 | 4 | 4 | 4 | 4 | 92 | 99 | 92.9 |
| GH 5 | 5 | 5 | 6 | 6 | 4 | 4 | 8 | 9 | 3 | 4 | 4 | 4 | 4 | 4 | 94 | 99 | 94.9 |
| RAD 3 | 4 | 5 | 6 | 6 | 4 | 4 | 8 | 9 | 4 | 4 | 4 | 4 | 4 | 4 | 97 | 99 | 98.0 |
| Percent agreement by evaluator | | 95.0 | | 91.7 | | 100.0 | | 91.7 | | 93.8 | | 100.0 | | 93.8 | | 95.7 | | † |

See notes at end of table.

**Table E-1. Science read-behind percent agreement, by scorer ID and science cluster: 2006—Continued**

| Science Cluster 3 | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 20 | 24 | 23 | 24 | 45 | 48 | 21 | 24 | 21 | 24 | 24 | 24 | 23 | 24 | 47 | 48 | 23 | 24 |
| MM 4 | 4 | 4 | 4 | 4 | 7 | 8 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 8 | 8 | 4 | 4 |
| ALG 2 | 3 | 4 | 4 | 4 | 7 | 8 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 8 | 8 | 3 | 4 |
| AIR 1 | 3 | 4 | 3 | 4 | 8 | 8 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 8 | 8 | 4 | 4 |
| AIR 3 | 4 | 4 | 4 | 4 | 8 | 8 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 7 | 8 | 4 | 4 |
| ED 4 | 3 | 4 | 4 | 4 | 8 | 8 | 2 | 4 | 3 | 4 | 4 | 4 | 3 | 4 | 8 | 8 | 4 | 4 |
| PM 7 | 3 | 4 | 4 | 4 | 7 | 8 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 8 | 8 | 4 | 4 |
| Percent agreement by evaluator | | 83.3 | | 95.8 | | 93.8 | | 87.5 | | 87.5 | | 100.0 | | 95.8 | | 97.9 | | 95.8 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 22 | 24 | 34 | 36 | 24 | 24 | 34 | 36 | 23 | 24 | 32 | 36 | 23 | 24 | 439 | 468 | 93.8 |
| MM 4 | 4 | 4 | 6 | 6 | 4 | 4 | 6 | 6 | 4 | 4 | 4 | 6 | 4 | 4 | 75 | 78 | 96.2 |
| ALG 2 | 2 | 4 | 6 | 6 | 4 | 4 | 5 | 6 | 4 | 4 | 6 | 6 | 3 | 4 | 70 | 78 | 89.7 |
| AIR 1 | 4 | 4 | 5 | 6 | 4 | 4 | 5 | 6 | 3 | 4 | 6 | 6 | 4 | 4 | 72 | 78 | 92.3 |
| AIR 3 | 4 | 4 | 6 | 6 | 4 | 4 | 6 | 6 | 4 | 4 | 6 | 6 | 4 | 4 | 76 | 78 | 97.4 |
| ED 4 | 4 | 4 | 6 | 6 | 4 | 4 | 6 | 6 | 4 | 4 | 6 | 6 | 4 | 4 | 73 | 78 | 93.6 |
| PM 7 | 4 | 4 | 5 | 6 | 4 | 4 | 6 | 6 | 4 | 4 | 4 | 6 | 4 | 4 | 73 | 78 | 93.6 |
| Percent agreement by evaluator | | 91.7 | | 94.4 | | 100.0 | | 94.4 | | 95.8 | | 88.9 | | 95.8 | | 93.8 | | † |

See notes at end of table.

**Table E-1. Science read-behind percent agreement, by scorer ID and science cluster: 2006—Continued**

| Science Cluster 4 | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 40 | 40 | 20 | 20 | 20 | 20 | 37 | 40 | 19 | 20 | 20 | 20 | 20 | 20 | 19 | 20 | 24 | 25 |
| MH 4 | 8 | 8 | 4 | 4 | 4 | 4 | 7 | 8 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 |
| M 1 | 8 | 8 | 4 | 4 | 4 | 4 | 8 | 8 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 5 | 5 |
| M 2 | 8 | 8 | 4 | 4 | 4 | 4 | 7 | 8 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 |
| WOG 3 | 8 | 8 | 4 | 4 | 4 | 4 | 7 | 8 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 |
| EF 6 | 8 | 8 | 4 | 4 | 4 | 4 | 8 | 8 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 |
| Percent agreement by evaluator | | 100.0 | | 100.0 | | 100.0 | | 92.5 | | 95.0 | | 100.0 | | 100.0 | | 95.0 | | 96.0 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 23 | 25 | 23 | 25 | 23 | 25 | 18 | 20 | 24 | 25 | 33 | 35 | 24 | 25 | 387 | 405 | 95.6 |
| MH 4 | 4 | 5 | 5 | 5 | 4 | 5 | 4 | 4 | 4 | 5 | 5 | 7 | 5 | 5 | 74 | 81 | 91.4 |
| M 1 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 4 | 5 | 5 | 7 | 7 | 5 | 5 | 79 | 81 | 97.5 |
| M 2 | 5 | 5 | 4 | 5 | 4 | 5 | 3 | 4 | 5 | 5 | 7 | 7 | 4 | 5 | 75 | 81 | 92.6 |
| WOG 3 | 5 | 5 | 4 | 5 | 5 | 5 | 4 | 4 | 5 | 5 | 7 | 7 | 5 | 5 | 79 | 81 | 97.5 |
| EF 6 | 4 | 5 | 5 | 5 | 5 | 5 | 4 | 4 | 5 | 5 | 7 | 7 | 5 | 5 | 80 | 81 | 98.8 |
| Percent agreement by evaluator | | 92.0 | | 92.0 | | 92.0 | | 90.0 | | 96.0 | | 94.3 | | 96.0 | | 95.6 | | † |

See notes at end of table.

**Table E-1. Science read-behind percent agreement, by scorer ID and science cluster: 2006—Continued**

| Science Cluster 5 | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 37 | 40 | 20 | 20 | 38 | 40 | 18 | 20 | 20 | 20 | 38 | 40 | 39 | 40 | 39 | 40 | 32 | 35 |
| GV2 | 7 | 8 | 4 | 4 | 7 | 8 | 3 | 4 | 4 | 4 | 8 | 8 | 8 | 8 | 7 | 8 | 5 | 7 |
| GV4 | 6 | 8 | 4 | 4 | 8 | 8 | 4 | 4 | 4 | 4 | 7 | 8 | 8 | 8 | 8 | 8 | 6 | 7 |
| DC1 | 8 | 8 | 4 | 4 | 7 | 8 | 4 | 4 | 4 | 4 | 7 | 8 | 7 | 8 | 8 | 8 | 7 | 7 |
| SS5 | 8 | 8 | 4 | 4 | 8 | 8 | 3 | 4 | 4 | 4 | 8 | 8 | 8 | 8 | 8 | 8 | 7 | 7 |
| PA6 | 8 | 8 | 4 | 4 | 8 | 8 | 4 | 4 | 4 | 4 | 8 | 8 | 8 | 8 | 8 | 8 | 7 | 7 |
| Percent agreement by evaluator | | 92.5 | | 100.0 | | 95.0 | | 90.0 | | 100.0 | | 95.0 | | 97.5 | | 97.5 | | 91.4 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 25 | 25 | 32 | 35 | 24 | 25 | 24 | 25 | 40 | 40 | 33 | 35 | 25 | 25 | 484 | 505 | 95.8 |
| GV2 | 5 | 5 | 7 | 7 | 5 | 5 | 5 | 5 | 8 | 8 | 5 | 7 | 5 | 5 | 93 | 101 | 92.1 |
| GV4 | 5 | 5 | 5 | 7 | 4 | 5 | 4 | 5 | 8 | 8 | 7 | 7 | 5 | 5 | 93 | 101 | 92.1 |
| DC1 | 5 | 5 | 6 | 7 | 5 | 5 | 5 | 5 | 8 | 8 | 7 | 7 | 5 | 5 | 97 | 101 | 96.0 |
| SS5 | 5 | 5 | 7 | 7 | 5 | 5 | 5 | 5 | 8 | 8 | 7 | 7 | 5 | 5 | 100 | 101 | 99.0 |
| PA6 | 5 | 5 | 7 | 7 | 5 | 5 | 5 | 5 | 8 | 8 | 7 | 7 | 5 | 5 | 101 | 101 | 100.0 |
| Percent agreement by evaluator | | 100.0 | | 91.4 | | 96.0 | | 96.0 | | 100.0 | | 94.3 | | 100.0 | | 95.8 | | † |

See notes at end of table.

**Table E-1. Science read-behind percent agreement, by scorer ID and science cluster: 2006—Continued**

| Science Cluster 6 | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 22 | 24 | 23 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 24 | 23 | 24 | 20 | 24 | 35 | 36 |
| IM1 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 6 |
| W1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 1 | 4 | 6 | 6 |
| W3a | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 6 | 6 |
| W3b | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 6 | 6 |
| BIM5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 6 | 6 |
| GP3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 6 | 6 |
| Percent agreement by evaluator | | 91.7 | | 95.8 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 95.8 | | 83.3 | | 97.2 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 24 | 24 | 18 | 18 | 35 | 36 | 23 | 24 | 24 | 24 | 29 | 30 | 30 | 30 | 402 | 414 | 97.1 |
| IM1 | 4 | 4 | 3 | 3 | 5 | 6 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 66 | 69 | 95.7 |
| W1 | 4 | 4 | 3 | 3 | 6 | 6 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 65 | 69 | 94.2 |
| W3a | 4 | 4 | 3 | 3 | 6 | 6 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 68 | 69 | 98.6 |
| W3b | 4 | 4 | 3 | 3 | 6 | 6 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 68 | 69 | 98.6 |
| BIM5 | 4 | 4 | 3 | 3 | 6 | 6 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 68 | 69 | 98.6 |
| GP3 | 4 | 4 | 3 | 3 | 6 | 6 | 3 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 67 | 69 | 97.1 |
| Percent agreement by evaluator | | 100.0 | | 100.0 | | 97.2 | | 95.8 | | 100.0 | | 96.7 | | 100.0 | | 97.1 | † |

See notes at end of table.

**Table E-1. Science read-behind percent agreement, by scorer ID and science cluster: 2006—Continued**

| Science Cluster 7 | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 39 | 40 | 39 | 40 | 40 | 40 | 39 | 40 | 39 | 40 | 40 | 40 | 39 | 40 | 40 | 40 | 40 | 40 |
| PE5 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 |
| TM1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| PI3 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| PI4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| DAD2 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| DAD3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| DAD4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| BAS1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| BAS2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| BAS3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Percent agreement by evaluator | | 97.5 | | 97.5 | | 100.0 | | 97.5 | | 97.5 | | 100.0 | | 97.5 | | 100.0 | | 100.0 |

| | 510 | | 511 | | 512 | | 513 | | 514[1] | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 39 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | † | † | 39 | 40 | 40 | 40 | 593 | 600 | 98.8 |
| PE5 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | † | † | 4 | 4 | 4 | 4 | 58 | 60 | 96.7 |
| TM1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | † | † | 4 | 4 | 4 | 4 | 60 | 60 | 100.0 |
| PI3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | † | † | 3 | 4 | 4 | 4 | 57 | 60 | 95.0 |
| PI4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | † | † | 4 | 4 | 4 | 4 | 59 | 60 | 98.3 |
| DAD2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | † | † | 4 | 4 | 4 | 4 | 59 | 60 | 98.3 |
| DAD3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | † | † | 4 | 4 | 4 | 4 | 60 | 60 | 100.0 |
| DAD4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | † | † | 4 | 4 | 4 | 4 | 60 | 60 | 100.0 |
| BAS1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | † | † | 4 | 4 | 4 | 4 | 60 | 60 | 100.0 |
| BAS2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | † | † | 4 | 4 | 4 | 4 | 60 | 60 | 100.0 |
| BAS3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | † | † | 4 | 4 | 4 | 4 | 60 | 60 | 100.0 |
| Percent agreement by evaluator | | 97.5 | | 100.0 | | 100.0 | | 100.0 | | † | | 97.5 | | 100.0 | | 98.8 | | † |

See notes at end of table.

**Table E-1. Science read-behind percent agreement, by scorer ID and science cluster: 2006—Continued**

| All science clusters | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 230 | 246 | 185 | 190 | 224 | 234 | 191 | 204 | 163 | 168 | 191 | 194 | 195 | 202 | 244 | 252 | 189 | 195 |
| Percent agreement by evaluator | | 93.5 | | 97.4 | | 95.7 | | 93.6 | | 97.0 | | 98.5 | | 96.5 | | 96.8 | | 96.9 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 166 | 173 | 184 | 193 | 176 | 181 | 201 | 215 | 141 | 146 | 196 | 206 | 171 | 174 | 3047 | 3173 |
| Percent agreement by evaluator | | 96.0 | | 95.3 | | 97.2 | | 93.5 | | 96.6 | | 95.1 | | 98.3 | | 96.0 |

† Not applicable.

[1] Scorer 514 did not complete Science Cluster 7.

NOTE: Read-behind percent agreement is the percentage a scorer and evaluator agree on a score given to a constructed response item. For example, if an evaluator checked 30 items by a Scorer ("Read" column) and agreed with the code given to open-ended responses for 26 of those items ("Agree" column), a percent agreement of 86.7 would result.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Program for International Student Assessment (PISA) 2006.

**Table E-2. Mathematics read-behind percent agreement, by scorer ID and mathematics cluster: 2006**

| Mathematics Cluster 1 | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 19 | 20 | 20 | 20 | 20 | 20 | 19 | 20 | 20 | 20 | 19 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| CD2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| CD3 | 3 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| H1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| MAB1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| B3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Percent agreement by evaluator | | 95.0 | | 100.0 | | 100.0 | | 95.0 | | 100.0 | | 95.0 | | 100.0 | | 100.0 | | 100.0 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 25 | 25 | 20 | 20 | 20 | 20 | 20 | 20 | 25 | 25 | 18 | 20 | 20 | 20 | 325 | 330 | 98.5 |
| CD2 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 4 | 4 | 4 | 4 | 66 | 66 | 100.0 |
| CD3 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 2 | 4 | 4 | 4 | 61 | 66 | 92.4 |
| H1 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 4 | 4 | 4 | 4 | 66 | 66 | 100.0 |
| MAB1 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 4 | 4 | 4 | 4 | 66 | 66 | 100.0 |
| B3 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 4 | 4 | 4 | 4 | 66 | 66 | 100.0 |
| Percent agreement by evaluator | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 90.0 | | 100.0 | | 98.5 | | † |

See notes at end of table.

**Table E-2. Mathematics read-behind percent agreement, by scorer ID and mathematics cluster: 2006—Continued**

| Mathematics Cluster 2 | Scorer ID 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 51 | 56 | 42 | 42 | 42 | 42 | 42 | 42 | 41 | 42 | 40 | 42 | 56 | 56 | 55 | 56 | 41 | 42 |
| RT1 | 7 | 8 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 8 | 8 | 8 | 8 | 6 | 6 |
| PP1 | 5 | 8 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 5 | 6 | 8 | 8 | 8 | 8 | 6 | 6 |
| PP2 | 7 | 8 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 8 | 8 | 8 | 8 | 6 | 6 |
| PP3 | 8 | 8 | 6 | 6 | 6 | 6 | 6 | 6 | 5 | 6 | 6 | 6 | 8 | 8 | 8 | 8 | 6 | 6 |
| D1 | 8 | 8 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 8 | 8 | 8 | 8 | 6 | 6 |
| B2 | 8 | 8 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 8 | 8 | 8 | 8 | 6 | 6 |
| TTS1 | 8 | 8 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 5 | 6 | 8 | 8 | 7 | 8 | 5 | 6 |
| Percent agreement by evaluator | | 91.1 | | 100.0 | | 100.0 | | 100.0 | | 97.6 | | 95.2 | | 100.0 | | 98.2 | | 97.6 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 41 | 42 | 39 | 42 | 41 | 42 | 40 | 42 | 55 | 56 | 55 | 56 | 42 | 42 | 723 | 742 | 97.4 |
| RT1 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 7 | 8 | 8 | 8 | 6 | 6 | 104 | 106 | 98.1 |
| PP1 | 6 | 6 | 5 | 6 | 6 | 6 | 6 | 6 | 8 | 8 | 8 | 8 | 6 | 6 | 101 | 106 | 95.3 |
| PP2 | 6 | 6 | 6 | 6 | 5 | 6 | 5 | 6 | 8 | 8 | 8 | 8 | 6 | 6 | 103 | 106 | 97.2 |
| PP3 | 6 | 6 | 5 | 6 | 6 | 6 | 5 | 6 | 8 | 8 | 8 | 8 | 6 | 6 | 103 | 106 | 97.2 |
| D1 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 8 | 8 | 8 | 8 | 6 | 6 | 106 | 106 | 100.0 |
| B2 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 8 | 8 | 8 | 8 | 6 | 6 | 106 | 106 | 100.0 |
| TTS1 | 5 | 6 | 5 | 6 | 6 | 6 | 6 | 6 | 8 | 8 | 7 | 8 | 6 | 6 | 100 | 106 | 94.3 |
| Percent agreement by evaluator | | 97.6 | | 92.9 | | 97.6 | | 95.2 | | 98.2 | | 98.2 | | 100.0 | | 97.4 | † |

See notes at end of table.

**Table E-2. Mathematics read-behind percent agreement, by scorer ID and mathematics cluster: 2006—Continued**

| Mathematics Cluster 3 | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 19 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 19 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| TTC1 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| TTC2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| CD1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| CD2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| CD3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Percent agreement by evaluator | | 95.0 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 95.0 | | 100.0 | | 100.0 | | 100.0 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 25 | 25 | 20 | 20 | 24 | 25 | 20 | 20 | 20 | 20 | 30 | 30 | 20 | 20 | 337 | 340 | 99.1 |
| TTC1 | 5 | 5 | 4 | 4 | 5 | 5 | 4 | 4 | 4 | 4 | 6 | 6 | 4 | 4 | 67 | 68 | 98.5 |
| TTC2 | 5 | 5 | 4 | 4 | 5 | 5 | 4 | 4 | 4 | 4 | 6 | 6 | 4 | 4 | 68 | 68 | 100.0 |
| CD1 | 5 | 5 | 4 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 6 | 6 | 4 | 4 | 66 | 68 | 97.1 |
| CD2 | 5 | 5 | 4 | 4 | 5 | 5 | 4 | 4 | 4 | 4 | 6 | 6 | 4 | 4 | 68 | 68 | 100.0 |
| CD3 | 5 | 5 | 4 | 4 | 5 | 5 | 4 | 4 | 4 | 4 | 6 | 6 | 4 | 4 | 68 | 68 | 100.0 |
| Percent agreement by evaluator | | 100.0 | | 100.0 | | 96.0 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 99.1 | | † |

See notes at end of table.

**Table E-2. Mathematics read-behind percent agreement, by scorer ID and mathematics cluster: 2006—Continued**

| Mathematics Cluster 4 | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| CW2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| RT1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| RT2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Percent agreement by evaluator | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 100.0 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | | Percent agreement by item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | |
| Total | 11 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 191 | 192 | 99.5 |
| CW2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 64 | 64 | 100.0 |
| RT1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 64 | 64 | 100.0 |
| RT2 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 63 | 64 | 98.4 |
| Percent agreement by evaluator | | 91.7 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 100.0 | | 99.5 | | † |

See notes at end of table.

**Table E-2. Mathematics read-behind percent agreement, by scorer ID and mathematics cluster: 2006—Continued**

| All mathematics clusters | Scorer ID | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 501 | | 502 | | 503 | | 504 | | 505 | | 506 | | 507 | | 508 | | 509 | |
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 101 | 108 | 94 | 94 | 94 | 94 | 93 | 94 | 93 | 94 | 90 | 94 | 108 | 108 | 107 | 108 | 93 | 94 |
| Percent agreement by evaluator | | 93.5 | | 100.0 | | 100.0 | | 98.9 | | 98.9 | | 95.7 | | 100.0 | | 99.1 | | 98.9 |

| | 510 | | 511 | | 512 | | 513 | | 514 | | 515 | | 516 | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number | | Number | | Number | | Number | | Number | | Number | | Number | | Number | |
| | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read | Agree | Read |
| Total | 102 | 104 | 91 | 94 | 97 | 99 | 92 | 94 | 112 | 113 | 115 | 118 | 94 | 94 | 1,576 | 1,604 |
| Percent agreement by evaluator | | 98.1 | | 96.8 | | 98.0 | | 97.9 | | 99.1 | | 97.5 | | 100.0 | | 98.3 |

† Not applicable.

NOTE: Read-behind percent agreement is the percentage a scorer and evaluator agree on a score given to a constructed response item. For example, if an evaluator checked 30 items by a Scorer ("Read" column) and agreed with the code given to open-ended responses for 26 of those items ("Agree" column), a percent agreement of 86.7 would result.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Program for International Student Assessment (PISA) 2006.

**Table E-3. Final percent agreement for constructed response items, by item cluster: 2006**

| Cluster | Unit name | Item ID | Percent agreement | Cluster | Unit name | ITEM ID | Percent agreement |
|---|---|---|---|---|---|---|---|
| M1 | Car Drive | M302Q02 | 99.09 | S2 | Greenhouse | S114Q03 | 85.96 |
| | Car Drive | M302Q03 | 91.82 | | Greenhouse | S114Q04 | 80.70 |
| | Height | M421Q01 | 92.73 | | Greenhouse | S114Q05 | 80.70 |
| | Bicycles | M810Q03 | 90.91 | | Radiotherapy | S495Q03 | 92.98 |
| M2 | Population Pyramids | M155Q02 | 81.82 | S3 | Mary Montagu | S477Q04 | 80.56 |
| | Population Pyramids | M155Q01 | 90.00 | | Algae | S268Q02 | 87.04 |
| | Population Pyramids | M155Q03 | 81.82 | | Airbags | S519Q01 | 77.78 |
| | Braille | M442Q02 | 95.45 | | Airbags | S519Q03 | 76.85 |
| | The Third Side | M462Q01 | 83.64 | | Experimental Digestion | S498Q04 | 82.41 |
| | | | | | Penicillin Manufacture | S524Q07 | 80.56 |
| M3 | The Thermometer Cricket | M446Q01 | 88.18 | | | | |
| | Carbon Dioxide | M828Q01 | 78.18 | S4 | Magnetic Hovertrain | S510Q04 | 94.44 |
| | Carbon Dioxide | M828Q02 | 93.64 | | Milk | S326Q01 | 80.56 |
| | Carbon Dioxide | M828Q03 | 95.45 | | Milk | S326Q02 | 79.63 |
| | | | | | Wild Oat Grass | S408Q03 | 82.41 |
| M4 | Running Tracks | M406Q01 | 92.73 | | Extinguishing Fires | S437Q06 | 87.96 |
| | Running Tracks | M406Q02 | 94.55 | | | | |
| | | | | S5 | Good Vibrations | S131Q02 | 86.24 |
| R1 | Employment | R219Q01E | 94.59 | | Good Vibrations | S131Q04 | 76.15 |
| | Employment | R219Q02 | 85.59 | | Different Climates | S465Q01 | 90.83 |
| | Aesop | R067Q04 | 87.39 | | Sunscreens | S447Q05 | 86.24 |
| | Aesop | R067Q05 | 80.18 | | | | |
| | Shirts | R102Q04A | 97.30 | S6 | The Ice Mummy | S458Q01 | 81.65 |
| | South Pole | R220Q01 | 92.79 | | Water | S304Q01 | 77.06 |
| | | | | | Water | S304Q03a | 82.57 |
| R2 | Optician | R227Q03 | 89.19 | | Water | S304Q03b | 87.16 |
| | Optician | R227Q06 | 97.30 | | Bacteria in Milk | S428Q05 | 84.40 |
| | Exchange | R111Q02B | 71.17 | | Green Parks | S438Q03 | 75.23 |
| | Exchange | R111Q06B | 79.28 | | | | |
| | Drugged Spiders | R055Q02 | 84.68 | S7 | Physical Exercise | S493Q05 | 88.99 |
| | Drugged Spiders | R055Q03 | 89.19 | | Penguin Island | S425Q03 | 91.74 |
| | Drugged Spiders | R055Q05 | 96.40 | | Penguin Island | S425Q04 | 86.24 |
| | Telephone | R104Q05 | 97.30 | | Development and Disaster | S514Q02 | 86.24 |
| S1 | Genetically Modified Crops | S508Q04 | 90.35 | | Development and Disaster | S514Q03 | 90.83 |
| | Earth's Temperature | S269Q01 | 92.11 | | | | |
| | Earth's Temperature | S269Q03 | 91.23 | | | | |
| | Grand Canyon | S426Q01 | 92.11 | | | | |
| | Acid Rain | S485Q02 | 88.60 | | | | |
| | Acid Rain | S485Q05 | 83.33 | | | | |