ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

User's Guide for the Progress in International Reading Literacy Study (PIRLS)

2006 Data Files and Database with United States Specific Variables

NCES 2009-050

U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

User's Guide for the Progress in International Reading Literacy Study (PIRLS)

2006 Data Files and Database with United States Specific Variables

March 2009

Patricia J. Green
Deborah Herget
Jeffrey Rosen
RTI International

Stephen Provasnik
Project Officer
National Center for Education Statistics

NCES 2009-050

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

March 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 Progress in International Reading Literacy Study (PIRLS): 2006 Data Files, and Database with United States Specific Variables* (NCES 2009-050). 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
Stephen Provasnik
(202) 502-7480
stephen.provasnik@ed.gov

Contents

List of Tables

Appendix Tables

Table **Page**

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

This user's guide is designed to provide researchers with an overview of the design and implementation of the Progress in International Reading Literacy Study (PIRLS 2006), as well as with information on how to access the PIRLS 2006 data. This information is meant to supplement that information presented in International Association for the Evaluation of Educational Achievement (IEA) publications by describing those aspects of PIRLS 2006 that are unique to the United States.

1.1 Overview of PIRLS 2006

PIRLS is a continuing assessment of the reading comprehension of students in their fourth year of schooling in jurisdictions around the world. PIRLS not only helps participating jurisdictions understand the literacy skills of their students but also places the literacy of young readers within an international context. Drawing comparisons between jurisdictions reveals areas of strengths as well as areas in need of improvement, offering jurisdictions insight into how the reading literacy of their students may be enhanced.

PIRLS is conducted by the International Association for the Evaluation of Educational Achievement (IEA), with national sponsors in each participating jurisdiction. In the United States, PIRLS is sponsored by the National Center for Education Statistics (NCES), in the Institute of Education Sciences in the U.S. Department of Education. NCES contracted with RTI International to conduct PIRLS 2006 in the United States.

PIRLS 2006 was the second cycle of the study, which was first administered in 2001. The assessment is open to countries and subnational entities. In this report, participating countries and subnational entities are both referred to as "jurisdictions." In 2006, 45 jurisdictions, including the United States, participated in PIRLS. In addition to 38 participating countries, this total includes 5 participating Canadian provinces and 2 separate samples of students that were assessed in Belgium.[1]

PIRLS provides trends and international comparisons on fourth-grade students' reading achievement, students' competencies in relation to goals and standards for reading education, and contexts of schooling, including student reports on reading attitudes and behaviors, as well as teacher and principal reports on key aspects of the schools, curriculum, and classrooms.[2]

PIRLS consists of a reading assessment and a set of questionnaires for students, their teachers, principals, and, as an international option in some countries, their parents (the United States did not participate in the PIRLS 2006 parent questionnaire). The kinds of questions that PIRLS data can address include:

- How does the reading literacy of fourth-grade students compare across countries?

[1] The two major geographic and cultural regions of Belgium (Flemish and French) have separate educational systems and were each assessed in PIRLS. Belgium (Flemish) and Belgium (French) are reported as separate jurisdictions

[2] PIRLS 2006 Assessment Framework and Specifications—2nd Edition, Mullis, I.V.S., Kennedy, A.M., Martin, M.O., & Sainsbury, M. (2006), Chestnut Hill, MA: Boston College

- How does the reading literacy of fourth-grade students vary by student background characteristics, school and classroom characteristics, and home reading environment? And how do these relationships vary across countries?
- How has reading literacy of fourth-grade students changed since 2001in the United States and other countries?

Further information about the technical aspects of the assessment is available in the international PIRLS 2006 Technical Report (Martin, Mullis, and Kennedy 2007). A comparison of the PIRLS and NAEP assessments of reading can be found in appendix A.

1.2 PIRLS 2006 Assessment Design

The PIRLS assessment used a Balanced Incomplete Block (BIB) design in which there were 13 versions of the test booklet. The design is premised on Item Response Theory which assumes that respondents are administered a sample of all items to produce aggregate group scores. As a result, no individual student was administered all the items. For the purposes of secondary analysis, each student record has five "plausible values" for each score that is estimated; however, it is important to keep in mind that these are provided solely so that data users can calculate scale scores for student groups defined by the analyst. In addition to having an elaborate assessment design, PIRLS used a complex sampling design. Consequently, analysts must use software and macros that have been specially developed for this type of analysis, as described in chapter 10.

1.3 How PIRLS 2006 Was Conducted in the United States

The process for conducting PIRLS in participating countries was overseen by the International Association for the Evaluation of Educational Achievement (IEA) which developed standards for all aspects of the study. The IEA oversight body communicated standards through various international meetings and a national research coordinator's manual. Each country was responsible for its own data collection, following the IEA standards. Quality monitors were hired by the IEA- funded PIRLS International Study Center at Boston College to observe the data collection and to ensure that guidelines were followed.

Conducting PIRLS 2006 in the United States involved sampling schools and classrooms (both teachers and students), recruiting schools and students to participate in the study, using the prescribed sampling software to record school participation and sample classrooms, developing and distributing instruments, collecting the PIRLS 2006 data, scoring constructed response items, and processing the data. These activities were conducted by RTI International under contract to NCES. Final cleaning and scaling of the data were completed for all countries by the IEA Data Processing Center (IEA-DPC) in Hamburg, Germany. Statistics Canada carried out all sampling and weighting tasks.

Chapter 2.
Sampling Schools and Students

2.1 PIRLS 2006 Sample Design

The PIRLS 2006 international project team instituted a series of sampling, data collection, and response rate benchmarks to ensure international comparability and to provide the ability to produce precise estimates of the main criterion variables for all jurisdictions.

The target population for PIRLS was defined by IEA using the International Standard Classification of Education (ISCED), developed by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) (UNESCO 1999). The target population of interest was all students enrolled in the grade corresponding to the fourth year of schooling, beginning with ISCED Level 1. For most jurisdictions, this was the fourth grade or its national equivalent. This definition is different from the one used in 2001, which targeted students in the upper of the two grades that include the most 9-year-olds, which in most jurisdictions was the fourth grade. Table 1 provides information on ISCED levels for the United States.

Table 1. International Standard Classification of Education (ISCED) levels, definitions, and U.S. equivalents in preprimary through 12th grade

ISCED level	Definition	U.S. equivalent
0	Preprimary	Kindergarten and below
1	Primary	1st–6th grades
2	Lower secondary	7th–9th grades
3	Upper secondary	10th–12th grades or first 3 years of vocational education

SOURCE: Matheson, N., Salganik, L., Phelps, R., Perie, M., Alsalam, N., and Smith, T. (1996). *Education Indicators: An International Perspective* (NCES 96-003). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Standardized procedures for sampling were developed by IEA and disseminated in a school sampling manual. Statistics Canada was responsible for approving the designs and verifying the samples of all participating jurisdictions. The basic sample design called for a two-stage stratified cluster design, with schools selected at the first stage and classrooms at the second stage. Schools were sampled using a probability proportionate to size sampling method. Within each jurisdiction a minimum of 150 schools were selected. Information on the number of classrooms containing fourth-grade students, and the size of the classes, were collected from participating schools and entered into the within-school sampling software provided by IEA. In most jurisdictions, one or two classes per school were randomly selected using this software. All students in sampled classrooms were selected.

IEA also established sample size and response rate targets for all jurisdictions. As table 2 shows, the weighted response rate target for schools was set at 85 percent, with a minimum participation rate among "original sample schools" of 50 percent. When the original sample was drawn, the schools immediately before and immediately after each sampled school on the sampling frame were designated "substitute" schools and were contacted if the original sample school refused to participate. The response rate target for classrooms was 95 percent, and the target weighted student response rate was set at 85 percent. In addition, classrooms with student

participation below 50 percent were to be rejected from inclusion with the final data. Substitution of sampled classrooms was not permitted, and the school would be classified as a nonrespondent if no other classrooms had been sampled. No U.S. schools were classified as nonrespondents on the basis of these criteria.

Table 2. IEA minimum sample size and weighted response rate targets for participating PIRLS jurisdictions: 2006

Group	Minimum sample size (number)	Weighted response rate (percent)
Schools	150	85[1]
Classrooms	1 per sampled school	95
Teachers	1 per sampled school	85
Students	4,500	85

[1] At least 50 percent must be original sample schools.
SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

The IEA's minimum acceptable rate for overall sample participation after substitution (the product of the weighted school participation rate and the weighted student participation rate) was 75 percent.

2.2 Sampling in the United States

The PIRLS sample in the United States was designed to be representative of all fourth-grade students in the 50 states and the District of Columbia. In addition to the base sample (designed to yield 150 participating schools), the United States sampled additional private schools and high-poverty schools, defined as those schools in which 50 percent or more of students were eligible to receive free or reduced-price lunch, to increase the precision of the estimates for these subgroups. The U.S. sample was designed to yield 180 participating schools.

The PIRLS school sample was drawn in March 2005. The sampling frame was constructed using data from the 2002–03 Common Core of Data (CCD) and preliminary 2003–04 data from the Private School Universe Survey (PSS).

To be consistent with the sampling design for PIRLS 2001, the frame was divided into two explicit strata: (1) one stratum was created that included schools located in the 10 most populous Metropolitan Statistical Areas (MSAs); (2) all schools outside those MSAs were grouped into 451 Primary Sampling Units (PSUs) by sorting on MSA and then by the Federal Information Processing Standards (FIPS) code. PSUs were designed to fit within state boundaries and, where possible, within county and city boundaries. In the United States, schools were sorted by state, percentage of racial/ethnic minority students, control of school (public/private), percentage of students eligible for free or reduced-priced lunch, and locale before the selection process.

Locale was determined on the basis of a sampled school's address. School addresses were classified into one of three categories (central city, urban fringe/large town, or rural/small town) using the NCES locale code system in use at the time of sampling. The locale code system used the following designations:

- Large city: A central city of a Consolidated Metropolitan Statistical Area (CMSA) or MSA, with the city having a population greater than or equal to 250,000.

- Midsize city: A central city of a CMSA or MSA, with the city having a population less than 250,000.
- Urban fringe of a large city: Any territory within a CMSA or MSA of a large city and defined as urban by the Census Bureau.
- Urban fringe of a midsize city: Any territory within a CMSA or MSA of a midsize city and defined as urban by the Census Bureau.
- Large town: An incorporated place or Census-designated place with a population greater than or equal to 25,000 and located outside a CMSA or MSA.
- Small town: An 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 CMSA or MSA.
- Rural, Outside MSA: Any territory designated as rural by the Census Bureau that is outside a CMSA or MSA of a large or midsize city.
- Rural, Inside MSA: Any territory designated as rural by the Census Bureau that is within a CMSA or MSA of a large or midsize city.

Large city and midsize city were combined to form central city; urban fringe of a large city, urban fringe of a midsize city, and large town were combined to form urban fringe/large town; and small town, rural, outside MSA, and rural, inside MSA were combined to form rural/small town.

Within each selected PSU or MSA stratum, schools were selected on the basis of the number of fourth-grade students in the school so that larger schools had a higher probability of selection than smaller schools. The final sample included 222 schools; 152 were chosen from PSUs and 70 were selected from the MSA stratum. In participating schools with at least two fourth-grade classes, two classrooms were randomly selected using specialized sampling software developed by IEA; in schools with one fourth-grade classroom, the one classroom was selected.

Chapter 3. Response Rates

3.1 Response Rates

Of the 222 sampled schools, 214 were eligible for inclusion in PIRLS, as shown in table 3. Eight schools had closed and were designated ineligible. Of the 214 eligible original sample schools, 120 participated (57 percent weighted). An additional 63 substitute schools agreed to participate, for a total of 183 schools, or a weighted response rate, using final adjusted weights, of 86 percent of eligible schools.[3] Of the 120 participating schools from the original sample, 88 (73 percent) were from the PSU sample, while 40 of the 63 participating substitute schools (63 percent) were from the PSU sample. The United States met the international guidelines for school response rate, but only after using substitute schools.

Information on the number and size of classrooms containing fourth-grade students was collected from all participating schools. One or two classrooms (depending on school size) were randomly selected from each school. Of the 256 classrooms sampled, 255 participated, or 99 percent.

In addition to having students complete the assessment and a questionnaire, PIRLS asked teachers and school administrators to complete questionnaires. Of the 255 teachers sampled, 249 completed teacher questionnaires, or 97 percent. Among school administrators, 182 of the 183 questionnaires were completed, for a response rate of 99 percent.

Table 3. School, administrator, and teacher response rates in the United States: PIRLS 2006

	Participating	Eligible	Weighted response rate (percent)
School participation before substitution	120	214	57
School participation after substitution	183	214	86
Classroom participation	249	256	99
School questionnaire completion	182	183	99
Teacher questionnaire completion	249	255	97

SOURCE: National Center for Education Statistics, Progress in International Reading Literacy Study (PIRLS), 2006.

Table 4 presents information on student response rates. There were 5,601 fourth-grade students enrolled in the selected classrooms; 159 of these students were excluded from testing (see "Exclusions" for more information). Therefore, within these classrooms, 5,442 students were eligible, and 5,190 completed the assessment, for an unweighted student response rate of 95 percent. Nine of the students who completed the assessment did not complete the questionnaire due to scheduling conflicts or illness, but are still considered to be completed cases. The United States met the international guidelines for classroom and student response rates.

[3] All weighted response rates discussed in this report refer to final adjusted weights. Response rates were calculated using the formula developed by the IEA for PIRLS. The standard NCES formula for computing response rates would result in a lower school response rate of approximately 63 percent.

Table 4. Student response rates in United States: PIRLS 2006

	Number	Percent
Total students sampled	5,761	100.0
Students withdrawn from class/school	160	2.7
Students excluded	159	2.7
Students eligible	5,442	94.5
Students absent or refused	252	4.6
Students assessed	5,190	95.4
Within-school student participation (weighted)		96.0

SOURCE: National Center for Education Statistics, Progress in International Reading Literacy Study (PIRLS), 2006.

Table 5 presents information on the total number of participating schools, students assessed, and overall weighted response rates after substitution in all participating jurisdictions.

Table 5. Total number of participating schools, students assessed, and overall weighted response rates, by participating PIRLS jurisdictions: 2006

Jurisdiction	Total number of participating schools	Total number of students assessed	Overall weighted response
Austria	158	5,067	97
Belgium (Flemish)	137	4,479	91
Belgium (French)	150	4,552	95
Bulgaria	143	3,863	94
Canada, Alberta	150	4,243	96
Canada, British Columbia	148	4,150	94
Canada, Nova Scotia	201	4,436	96
Canada, Ontario	180	3,988	87
Canada, Quebec	185	3,748	81
Chinese Taipei	150	4,589	99
Denmark	145	4,001	96
England	148	4,036	92
France	169	4,404	95
Georgia	149	4,402	98
Germany	405	7,899	92
Hong Kong, SAR	144	4,712	97
Hungary	149	4,068	97
Iceland	128	3,673	90
Indonesia	168	4,774	98
Iran	236	5,411	99
Israel	149	3,908	93
Italy	150	3,581	97
Kuwait	149	3,958	88
Latvia	147	4,162	92
Lithuania	146	4,701	92
Luxembourg	178	5,101	99
Macedonia	150	4,002	96
Moldova	150	4,036	95
Morocco	159	3,249	94
Netherlands	139	4,156	90
New Zealand	243	6,256	95
Norway	135	3,837	71

See notes at end of table.

Table 5. Total number of participating schools, students assessed, and overall weighted response rates, by participating PIRLS jurisdictions: 2006—Continued

Jurisdiction	Total number of participating schools	Total number of students assessed	Overall weighted response
Poland	148	4,854	95
Qatar	119	6,680	94
Romania	146	4,273	97
Russian Federation	232	4,720	97
Scotland	130	3,775	81
Singapore	178	6,390	95
Slovak Republic	167	5,380	94
Slovenia	145	5,337	93
South Africa	397	14,657	88
Spain	152	4,094	97
Sweden	147	4,394	96
Trinidad and Tobago	147	3,951	94
United States	183	5,190	82

NOTE: The overall weighted response rate is the product of the school participation rate, after substitution, and the student participation rate.
SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

3.1.1 Exclusions

Schools that were very small or that were classified as special education, vocational, or alternative schools (private and public) could be excluded from the sampling frame. In the United States these schools enrolled 3.2 percent of the expected number of fourth-grade students. Table 6 shows the percentage of students excluded from the sample in 2001 and 2006.

Table 6. Percentage of U.S. students excluded from PIRLS at the school-listing level and student-listing level: 2001 and 2006

Level	2001	2006
Total	5.3	5.9
Excluded at the school listing level	0.6	3.2
Excluded at the student listing level	4.7	2.8

NOTE: Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2001 and 2006.

International guidelines recognized that some students might not be eligible for inclusion in PIRLS because of limited exposure to the language of assessment (English in the case of the United States) or the need for special testing accommodations.

Within classrooms, students were excluded from participation in PIRLS if they met the criteria established by the IEA:

- *Students with functional disabilities.* These are students who are permanently physically disabled in such a way that they cannot perform in the PIRLS testing situation. Functionally disabled students who could perform were included in the testing.

- *Students with intellectual disabilities.* These are students who are considered in the professional opinion of the school administrator or by other qualified staff members to be intellectually disabled or who have been psychologically tested as such. This includes students who are emotionally or mentally unable to follow the general instructions of the test. Students were not excluded solely because of poor academic performance or normal disciplinary problems.
- *Non-native language speakers.* These are students who are unable to read or speak the language of the test and would be unable to overcome the language barrier in the test situation. Typically, students who received less than 1 year of instruction in the language of the test were to be excluded, but this definition could be adapted in different jurisdictions. In the United States, English was the exclusive language of the assessment and students who had received less than 1 year of English instruction were defined as non-native language speakers.

In the United States, 2.8 percent of students in classrooms sampled for PIRLS were excluded on the basis of these criteria. In keeping with international protocol, no testing accommodations were offered to students.

The overall exclusion rate was 5.9 percent in the United States, which means that the overall U.S. coverage rate is .09 percent below the recommended 95 percent. Other jurisdictions that had exclusion rates above 5.0 percent included Bulgaria (6.4); the province of Ontario, Canada (8.3); Israel (22.5); Italy (5.3); Lithuania (5.1); New Zealand (5.3); and the Russian Federation (7.7).

3.1.2 Nonresponse Bias Analysis

A nonresponse bias analysis was conducted because the study did not achieve an 85% response rate among original sample schools, as required by the NCES statistical standards (the complete report can be found in appendix B.) The analysis of school nonresponse was conducted in two parts. The basis for both analyses was the original sample of 214 eligible schools. First, the distribution of the 120 responding original sample schools was compared with that of the total sample of eligible original schools. All original schools in the sample that declined to participate in the study were treated as nonparticipants regardless of whether they were replaced by a substitute school. In the second part, substitute schools were included in the analysis, reflecting the final sample of schools that participated in PIRLS 2006.

Seven variables were examined using the original sample, the participating schools from the original sample, and the participating schools in the final sample: (1) public/private school control, (2) locale, (3) region, (4) percentage of students eligible for free or reduced-price lunch, (5) total school enrollment, (6) fourth-grade enrollment, and (7) relative enrollment of racial and ethnic groups (White, non-Hispanic; Black, non-Hispanic; Hispanic; Asian or Pacific Islander; American Indian or Alaska Native; and other).

Measures of bias and relative bias were computed, and the hypothesis of independence between the characteristic and participation status was tested using chi-square statistics. In addition, logistic regression models were used to evaluate whether any of these characteristics were significant in predicting response status.

For the original sample there is evidence of some significant nonresponse bias before weight adjustment with respect to total enrollment and the race/ethnicity composition of the responding schools, especially the Asian, Hispanic, and Other categories. However, after weight adjustment, the significant bias is limited to the Asian and Other students with biases of 0.9 and -3.0 percent, respectively. In addition, the relative bias before weight adjustment indicates the potential for bias in school control, community type, Census region, and poverty level. After weight adjustment, the relative bias is about 14 percent for the Northeast Census region and about 27 percent for the Central Census region.

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 with respect to race/ethnicity, school control (private/public), and enrollment. However, after weight adjustment, no significant bias remains. In addition, the relative bias before weight adjustment indicates the potential for bias in community type, Census region, poverty level, and free or reduced-price lunch status. After weight adjustment, the relative bias is about 12 percent for private schools.

Overall, the data suggest that while significant nonresponse bias and large relative biases exist before weight adjustments, the biases are generally reduced after weight adjustment. 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.

Chapter 4.
Recruitment of Schools and Students

4.1 Task Force

After experiencing difficulties achieving high levels of school participation in recent international studies, NCES and the Education Statistics Services Institute (ESSI)[4] convened a task force prior to the start of PIRLS 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 PIRLS 2006 recruitment effort, the following recommendations were followed:

- Begin the recruitment of schools for PIRLS at least a year prior to the scheduled data collection.
- Approach selected schools for participation in PIRLS 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.

4.2 Institution Contacting

The eligible sample consisted of 214 original schools from 168 districts. The plan for recruiting schools incorporated the recommendations of a task force convened by NCES to improve school participation rates on international studies, including the following:

- School contacting began a year in advance of the scheduled data collection to try to reduce scheduling conflicts and allow time for schools' internal review and clearance processes.
- Sampled schools were contacted directly after an informational letter was sent to states and public school districts alerting them that schools in their state/district had been selected; and schools that refused were scheduled for in-person contacts when appropriate.

School contacting began on April 18, 2005, a few weeks after the final sample was received from Statistics Canada. The data collection was originally scheduled to take place in March and April 2006. During the recruitment process, many schools were reluctant to commit to the assessment because of mandatory state testing in March and April. Thus, the United States obtained permission to extend the data collection window from January 28 through June 9, 2006, to enable more schools to work PIRLS into their calendars.[5]

[4] ESSI provides developmental, analytical, methodological, and operational support to the National Centers for Education Statistics (NCES), Institute of Education Sciences, and U.S. Department of Education.

[5] This resulted in a 4½-month test window rather than the more typical 1- to 2-month test window. The mean score of students completing the examination in January through March was 539.5, which was not significantly different from the score (541.1) of the students completing the examination in April through June.

Prior to the field test, endorsements were secured from 14 organizations for PIRLS 2006. Copies of the endorsement letters were included in materials sent to schools. See appendix C for a list of the 14 organizations that endorsed PIRLS 2006.

4.2.1 State and District Notification

On April 18, 2005, informational materials were mailed to each of the 34 Chief State School Officers (CSSOs) from states selected for the PIRLS 2006 main study. The package was addressed to the CSSO and contained a lead letter from the Associate Commissioner of NCES, a study fact sheet, a sample copy of our implied (passive) parental consent form, and copies of the endorsement letters. The letter to the CSSO included a toll-free number to call if the state official had any questions. Copies of the packet were sent to the State Assessment Coordinator for each state. All state-level packages were sent by U.S. Mail.

Also on April 18, 2005, informational materials were mailed to each of the 168 districts or dioceses that had schools in the original sample. The package was addressed to the District Superintendent and contained the same materials as included in the state mailing. The letter to the Superintendent indicated that we would be contacting the sampled schools in the next few days. See appendix D for copies of the recruitment materials.

Letters to districts with only original schools indicated the sampled schools that were selected to participate. For districts with both original and substitute schools, the letter stated that the original school would be contacted and that the substitute school(s) might be contacted if the originally sampled school was unable to participate.

If a substitute school was needed because an original school refused to participate, district letters were then sent on a flow basis to the districts with only one or more substitute schools in the sample. The district superintendents received a package that was identical to the one received by superintendents of districts with original schools. In total, 167 school districts were sent informational mailings.

There were a handful of districts for which it was known, based on prior experience, that a formal application and district approval procedures would need to be followed. In those situations, the necessary documentation was included in the initial mailing to the district. There were several other districts that were also found to require application or approval procedures. In total, there were 5 districts that required applications and 17 that had other processes that were necessary before contacting the school in those districts.

4.2.2 School Contacting

Within a few days of sending the notification materials to the states and districts, each sampled original school was sent an informational package by Federal Express. The package was addressed to the principal and contained a lead letter from the NCES Associate Commissioner for the, Early Childhood, International, and Crosscutting Studies Division, National Center for Education Statistics. It also included a study fact sheet, a sample copy of the implied (passive) parental consent form, and copies of the endorsement letters.

Several days after the package was sent, the school was contacted by telephone. After determining the appropriate person with whom to speak, the recruiter discussed details about the study and answered any questions. If the school agreed to participate, the principal was asked to identify a school coordinator. This person served as a point of contact at the school and was

responsible for handling the logistical arrangements. School coordinators were offered an honorarium for their efforts to make the studies successful at each school. The coordinator received $75 cash at the end of the initial test session. The recruiter also scheduled a date for test day and date as a makeup day.

Several conversion strategies were employed with schools that declined to participate. For example, project staff reviewed case notes frequently and reassigned cases to survey refusal specialists as needed. In addition, at the recommendation of the task force, in-person visits were scheduled with schools initially refusing to participate. Of the 21 in-person conversion visits conducted, six were successful. Mailings were also sent to all schools near the end of the recruiting period, asking them to reconsider participation and providing information about the medals and certificates that participating students would be given.

4.2.3 Procedures Used for Contacting

Since many of the contacts with schools would be made by phone, a team of telephone recruiters was formed who were responsible for contacting and recruiting districts and schools for PIRLS 2006. The recruiting team consisted of institutional contactors from the call center, experienced field staff, and project survey specialists.

During a 1-day training, the overall project objectives and study components were discussed so that contactors would be able to answer questions. The schedule of mailings and other activities already completed were also presented. Finally, institutional contactors were given time to practice using the computerized institutional contacting system (ICS). The ICS was used to track all contacts with schools and was also used to generate reports to track progress and prepare materials for mailings.

4.2.4 Coordinator Manual Mailing/Within-School Sampling

The first request made of the school coordinator was for him or her to provide the number of classrooms containing fourth-grade students and the number of students within each of those classes. Each classroom was identified by the teacher's name or class number. Most of the schools chose to provide the information to our institutional contactors by phone.

Once the list of classrooms was collected, the information was keyed into WinW3S, the sampling software provided by the international consortium. Classroom sampling was then performed by WinW3S. One or two classrooms per school were selected by the software program.

Schools were notified which classroom(s) was (were) sampled, and schools were asked to provide information on each student in the sampled classroom(s). Once the student-level information was received, it was entered into WinW3S and test booklets were assigned to each student, using the booklet rotation algorithm programmed into the software. Schools were not asked to indicate students who should be excluded from testing until closer to the data collection period (this was done just prior to test day when the test administrator could personally discuss the exclusion guidelines with the coordinator), so all students were assigned a test booklet. Student tracking forms (STFs) and booklet labels were generated by WinW3S and available as PDF files once booklet assignment was complete. Copies of student tracking and other forms can be found in the School Coordinator Manual (TIMSS & PIRLS International Study Center 2005).

4.3 Securing Parental Permission

During the recruitment process, options for the parental permission process were discussed with the schools. Schools were encouraged to select passive (or implicit) parental consent unless the school expressed the need for active (or explicit) consent. A total of 21 schools (11 percent) required active parental consent and the remaining 162 schools used passive consent.

4.3.1 Active Consent

For schools that required active parental consent, packets were sent to the school coordinators to be distributed to the students selected to take the test. The packets contained a letter about the study, a consent form, information about the study, and an envelope bearing the teacher's name so parents could return the signed consent form. The consent forms were sent about 2 weeks prior to each school's scheduled test day. Prior to test day, the test administrators checked with the coordinators to obtain the names of parents who had not yet sent back a consent form. If they were given telephone numbers, the test administrators telephoned the parents to prompt them to return the forms. The student response rate in schools that required active consent was 92.2 percent.

4.3.2 Passive Consent

The majority of the schools used passive parental consent forms. Packets with consent materials were sent to the school coordinators to be distributed to the students selected to take the test. The packets contained a letter about the study, a consent form, and information about the study. Test administrators contacted the school coordinators prior to test day to determine if any parents had sent back forms that refused consent. For those parents, the test administrators attempted refusal conversion if the school was willing to provide telephone numbers. The student response rate in schools that required passive consent was 96 percent.

Chapter 5.
Instrument Development and Distribution

5.1 Test Development

The International Study Center (ISC), which organized and managed the international components of PIRLS, developed an assessment framework used to guide the test development process (Mullis et al. 2007). PIRLS was designed to assess two purposes of reading: reading for literary experience and reading to acquire and use information. In addition, the PIRLS assessment evaluates four processes of comprehension: (1) to focus on and retrieve explicitly stated information; (2) to make straightforward inferences; (3) to interpret and integrate ideas and information; and (4) to examine and evaluate content, language, and textual elements.

Jurisdictions participating in PIRLS 2006 were invited to submit reading passages to be used in the test. Two types of passages were sought: literary texts, which were typically narrative fiction, and informational texts, which could include biographies, step-by-step directions, informational leaflets, and scientific or other nonfiction material. All passages were to be authentic texts typical of the reading material in their jurisdictions, well suited to fourth-grade students, and no longer than 1,000 words. The national research coordinators from participating jurisdictions were asked to review the texts and work together to agree on a shortened list of passages to be illustrated and formatted. Questions for each passage were refined by PIRLS International Study Center staff and reviewed by a group of reading experts. Each reading passage, including text and questions, was designed to be completed in 40 minutes.

Through this process, twelve new passages were created and tested during a field trial in spring 2005. Item statistics, including item difficulties, point biserial correlations, and item discrimination statistics, were calculated for each item for each jurisdiction. After a careful review of the quality of all items across jurisdictions, six of these passages, three literary and three informational, were selected for the main study.

These passages, along with four passages from PIRLS 2001, were used to create the test booklets for the main study. The same 10 passages were used in all participating PIRLS jurisdictions. Each test booklet contained two reading passages. Students were given 40 minutes to complete each passage, or 80 minutes in all. The passages were distributed across 13 booklet types. Students were asked to answer a number of items related to each passage, including both multiple-choice and constructed-response items. The distribution of the items by type of passage and type of item is shown in table 7.

In addition to the assessment, students were asked to complete a 20- to 30-minute questionnaire. The questionnaire included items about students' reading experiences in school, self-perception and attitudes toward reading, out-of-school reading habits and computer use, home literacy resources, and basic demographic information.

Table 7. Distribution of items on the PIRLS 2006 assessment

Reading purpose	Multiple-choice items	Constructed-response items			Total number of items	Total score points
		1 point	2 points	3 points		
Total	64	28	27	7	126	167
Literary	34	13	13	4	64	85
Informational	30	15	14	3	62	82

SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

5.2 Translation and Adaptation

Source versions of all instruments (assessment booklets, questionnaires, and manuals) were prepared in English and then translated into the primary language or languages of instruction in each jurisdiction. In addition, it was sometimes necessary to adapt the instrument for cultural purposes, even in jurisdictions such as the United States that use English as the primary language of instruction. For example, words such as "lift" might be adapted to "elevator" for the United States. The IEA and ISC verified the translations and adaptations used by all participating jurisdictions. Certified translators were retained by the IEA to compare national versions with the source versions of all documents.

Chapter 6.
Overview of Field Operations

6.1 Recruitment of Field Supervisors and Test Administrators

The data collection contractor hired 4 Field Supervisors (FSs) and 38 Test Administrators (TAs) to work on the PIRLS 2006 main study. The FSs had experience working on several education and other research studies. Fifteen of the TAs had worked on school-based studies within the past year and the other TAs had experience on a variety of other research studies.

6.2 Training of Field Supervisors and Test Administrators

Prior to training, each TA received a copy of the test administrator manual 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 Durham, NC, on January 5–6, 2006. The FSs attended a 1-day training session the day prior to the TA training. Each FS and TA signed a confidentiality agreement and an affidavit of nondisclosure at the beginning of training.

The purpose of the Field Supervisor training was to prepare the FSs for the subsequent training. The FSs received an overview of PIRLS and were given time to practice completing the required forms. They also received instructions on supervising the TAs, monitoring data quality and production costs, and using the Field Reporting System to record and monitor data collection outcomes.

During the TA training, project staff discussed contacts that had already been made with the schools, as well as contacts that each TA would need to make with the school coordinator prior to test day. The training covered test day logistics and administration instructions for the student tests and questionnaire, including the international protocols and administrative procedures. TAs were trained to understand a school logistics form containing information about the school, such as the name and phone number of the school coordinator and the designated test day and makeup day. Trainers discussed criteria for scheduling makeup sessions and how to set them up with the coordinator. While explaining active and passive consent procedures, the group discussed contacting parents for gaining active permission and converting refusals. As requested during a TA debriefing after the field test data collection, several hands-on exercises were used to help the TAs better understand how to categorize students with exclusion codes and how to code a student who was deemed unable to participate by the school but was able to participate according to international standards. For example, many students with learning disabilities were considered by schools to be exclusions but were eligible according to the international guidelines; these were labeled refusals rather than exclusions.

At the end of the training, the TAs were tested to ensure that they were prepared to conduct PIRLS 2006 administrations in the schools. The TAs were certified on their ability to answer frequently asked questions about PIRLS, read the student script, understand the school logistics form, and complete the student tracking form. Each of the TAs successfully completed

certification, although there were concerns expressed about the TAs' readiness to complete the student tracking forms. Thus, additional STF exercises were developed and sent to the TAs to complete and return to supervisors to be reviewed prior to conducting any test administrations.

6.3 Description of Student Test Day and Makeup Sessions

About 2 weeks prior to the scheduled administration, TAs received student tracking forms for each of his or her assigned schools. At the same time, school coordinators were sent a packet of materials. The coordinator's packet included the student tracking forms as well as parental consent forms, reminders for the students about the approaching test date, and school and teacher questionnaires. Coordinators were asked to distribute the questionnaires to the designated respondents.

Prior to the designated test day, the TA phoned the coordinator. The purpose of the contact was to make sure the consent forms arrived and were distributed, to confirm the logistical arrangements for the session, and to verify that the school and teacher questionnaires had been distributed to the staff. The TA also reviewed the student tracking form with the coordinator to discuss any students who needed special accommodations and to determine whether they would be able to participate. At the same time, the TA determined if the coordinator had received any parental refusal forms. If so, the TA began refusal conversion efforts provided the school was willing to provide a telephone number for the parent. In active consent schools, the TA also learned from the coordinator which parents had not yet returned permission forms. If the school was willing to provide telephone numbers, the TA began calling the parents to prompt them to return the forms.

The test booklets and student questionnaires were sent directly to the TA at least 1 week prior to the scheduled session. Each test booklet and questionnaire was labeled prior to sending the administration material to the TA. The TA brought the materials with him or her to the school. The TAs arrived at each school about 1 hour prior to the scheduled administration to set up for the session.

On the test day at each school, the TA checked in with the school coordinator and collected any parental permission forms that had come in. In active consent schools, the TA checked the student tracking form to make sure that only students who had returned signed permission forms were allowed to participate. In both active and passive consent 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 TA distributed the labeled test booklets, questionnaires, and pencils to each student. Prior to beginning data collection, the TA read a script to the students describing the study, giving the elements of informed consent and giving instructions for completing the test and questionnaire.

The fourth-grade students were randomly assigned 1 of 13 test booklets. There was one version of the student questionnaire for PIRLS. The student portion of PIRLS took 2 hours in total. The test consisted of two 40-minute blocks with a 5-minute break in between. Students received another 5-minute break after completing the test before taking the 20- to 30-minute student questionnaire. When all students had finished, the TA distributed a book list with the titles of 10 books that students could choose from for their completion award. Students were asked to indicate on the book list their first, second, and third choice. One book for each child

was sent to the school within 4 weeks of the session. The TA also gave each student a medal and a certificate stating that the student represented the United States in PIRLS 2006.

At the conclusion of the session, the TA determined whether a makeup day was necessary. For PIRLS, a makeup session was held for each classroom that did not achieve at least 90 percent student participation. If a makeup day was deemed necessary, the TA informed the school coordinator. Makeup days were conducted at 38 schools. When makeup sessions were required in a school with two classrooms of testing, a joint session was held with the students from both of the selected classes that missed the test day.

Once the TA returned home from a test administration, he or she contacted the FS to report on the outcomes of the session. The participation status of each student was recorded into the Field Reporting System (FRS). This information was used to track the materials that were expected against those that were actually received. Reports were generated from the FRS to identify any discrepancies between what the FS recorded and what was sent in by the TA. The primary types of discrepancies the report would indicate were missing booklets, receiving a booklet for a student who was recorded as a nonparticipant, or receiving a refusal form 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.

6.4 School and Teacher Questionnaires

In addition to the student component of PIRLS, there were also school and teacher questionnaires. A school administrator from each participating school was asked to complete the School Questionnaire, and the primary teacher for each selected classroom was asked to complete a Teacher Questionnaire. It was preferred that the principal complete the school questionnaire, but any knowledgeable school staff member could complete the majority of the questionnaire. If more than one teacher was responsible for teaching the students in the sampled classroom, the primary teacher was asked to complete the questionnaire, consulting with other teachers as needed.

An envelope was included in the consent mailings to the school coordinator to be distributed to the designated respondent for the school questionnaire, and another was included for each teacher whose fourth-grade classroom was selected for PIRLS 2006. We asked the coordinator to distribute the envelopes to the designated respondents, as noted on the label on the front of the envelope. The envelope contained a letter from NCES and a questionnaire on which a label was affixed with the teacher name and classroom information.

School administrators and teachers were asked to complete the questionnaire at their convenience between the time of its receipt and the time that the TA came to conduct the session. The TA collected the completed forms on test day, and if the questionnaire was lost or not yet completed, the TA provided a second form and asked the school administrator or teacher to complete it while the test was in session. Most of the forms were completed before or during the student test administration. For school administrators or teachers 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 the data collection contractor prompted nonrespondents by phone to complete and return the forms.

6.5 Monitoring Data Collection Progress

Several reports were run daily to monitor the progress of the PIRLS 2006 data collection. The data collection team used some of the reports to help with the preparations for test day at each school. In particular, the Getting Ready for test day report was reviewed to ensure that all of the logistical information was obtained in time to send the materials to the TA and the school. A school participation report allowed data collection staff to monitor changes in school participation.

Once data collection was completed at a school, other reports were used to help the data collection team monitor progress. A response rate report provided response rate information for each participating school as well as a breakdown of reasons for nonparticipation. Another report compared the data that were received at RTI against the information recorded in the Field Reporting System (FRS), and provided timely information about issues that needed to be followed up on or corrected. Reports were also used to track the number of each booklet type completed, and response rates for school and teacher questionnaires.

After the schools completed the administration, the institutional contactors conducted verification re-interviews with the school coordinators and principals to verify that the administration went well. Information about the session was entered into the FRS and feedback was provided to the TAs.

Chapter 7.
Scoring and Coding

7.1 Prior to Training Scorers

CTB-McGraw Hill conducted the PIRLS 2006 scoring in the United States. Prior to training the U.S. scorers, supervisors from CTB completed the international scoring training sessions and devised a detailed organizational system for scoring. Two scoring supervisors from CTB attended the international training in Luxembourg, March 20–23, 2006, and were responsible for training the CTB scoring staff and monitoring the quality of the scoring process.

The PIRLS team consisted of a team leader and six scorers. As required, all employees signed a notarized confidentiality agreement. The staff then read the reading passages, answered test questions, previewed the scoring guides, and reviewed a selection of the anchor papers. At this point, scorers were ready to train on each PIRLS item.

7.2 Training

Scorers were trained on each item for a given story as follows. First, a scoring supervisor read aloud the scoring guide. Next, anchor papers were read aloud and discussed. Score rationales included in the scoring guide were used to help clarify and answer questions when needed. Next, each scorer worked through the practice papers individually, recording their scores on a form supplied by CTB. The forms were collected and individual scores recorded to make sure an acceptable level of understanding was maintained. After redistributing the practice paper scoring forms to the scorers, each practice response was read aloud by a scorer and there was group discussion if there were questions or disagreement about the correct score. Scorers were encouraged to take turns explaining their scoring rationale to the group. All scoring questions were thoroughly discussed before proceeding to the next item. All items for a given passage were scored in all booklets before training on the next passage commenced.

7.3 Trend Scoring During Training

Four passages were used in both PIRLS 2001 and PIRLS 2006. To ensure that scoring was conducted consistently across years, IEA developed a trend scoring exercise. CTB was supplied with a sample of student booklets scored in 2001. The Team Leader and each of the scorers participated in the trend scoring. After reading one of the 2001 responses, a scorer entered the score into specially developed software that compared that score to the one originally assigned to the response in 2001. The software reported agreement of each scorer with 2001. An 85 percent or better rate of agreement between the 2001 and 2006 scorers was required to score the 2006 trend items. Approximately half of the 200 or so student responses were scored for each item before scoring of 2006 items was allowed to begin.

The remaining trend scoring was completed during the scoring of the 2006 student booklets. If the agreement rates were below 85 percent at any point, re-training was performed and the trend scoring for that item was done again until the standards were met. The final overall rate of exact agreement between 2001 and 2006 scores was 93 percent in the United States, and

90 percent exact agreement on average across countries (Mullis, Martin, Kennedy and Foy 2007: 301).

7.4 Scoring

Test booklets were arranged in stacks ranging from a single booklet to approximately 10 booklets. Each stack of booklets represented a batch and had a unique batch number. The scorers were divided into two groups. Most of the batches contained one or more booklets marked for reliability scoring. That is, booklets that would be scored by two different scorers. After training, the batches of booklets were distributed to the two groups along with reliability scoring sheets for recording scores of the booklets selected to be reliability booklets. The reliability sheets were used so that the second scorer would be unable to see the scores assigned by the initial scorer. After the reliability booklets were scored, the booklets were returned for redistribution to the other group. Scorers would then score complete batches of booklets. A tally sheet was available to make sure that each scorer was scoring reliability booklets from a variety of scorers in the other group. This process was repeated for each passage.

7.5 Quality Monitoring

Throughout the scoring process, team leaders rescored booklets that had been scored by people in their groups. (Appendix E contains information on scorer agreement across items). If the team leader questioned a score given by a scorer, he or she either discussed the item with the individual scorer or discussed it with the group. In some cases, the item had been scored incorrectly, and the score would be changed. In other cases the response did not clearly fit into the training material and a decision would be made. In most cases a consensus was reached and agreed upon as to how to score similar responses should they arise again. As a further quality control check, the scoring supervisor also scored booklets that had been scored by the team leader to ensure that the team leader was continuing to score accurately.

The PIRLS IEA-DPC also conducting scoring reliability analyses for all countries after data collection and scoring were completed. The rate of exact agreement between scorers in the United States was 93 percent, the same as the international average. The international report (Martin, Mullis, Kennedy and Foy 2007) contains a detailed analysis of scorer reliability across countries in appendix A of that report.

Chapter 8.
Editing, Processing, and Weighting

8.1 Data File Cleaning and Editing

Within 48 hours of each test administration, the Test Administrator (TA) sent the completed materials to the data collection contractor for processing. 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.

A computerized system was developed to record the receipt of each form at the student level, including the test booklet, questionnaire, and consent form (if applicable). Once the items were received and recorded in the system, they were batched in groups of 10 and stored until data entry or scoring.

Four data entry staff persons were trained to key the test booklets, questionnaire, and reliability sheets. In addition, a data entry supervisor was trained to do the 5 percent double entry of each set of materials. The data entry staff keyed the data directly into WinDEM.

Once all of the booklets were keyed, the prescribed validation reports developed by the IEA-DPC were run to identify inconsistencies in the database and potential problems. Cleaning the data consisted of reviewing the student tracking forms, test administration forms, test booklets, and questionnaire booklets to resolve discrepancies that appeared on the reports. Staff cleaned as many discrepancies as possible and annotated the remaining issues, since at times items identified as problematic on the reports were actually an accurate account of the test administration at the school.

The data were submitted by the United States to the IEA-DPC in early September 2006. The IEA then conducted another stage of cleaning for all countries, notifying the national centers of any issues encountered. Following a few queries by the IEA-DPC, the U.S. data were swiftly finalized.

8.2 Weighting and Variance Estimation

Using sampling weights is necessary for computing statistically sound, nationally representative estimates. Survey weights help adjust for the international over- or under-sampling of certain sectors of the population, school or student nonresponse, or errors in estimating the size of a school at the time of sampling. Survey weighting for the entire international PIRLS 2006 sample was carried out by Statistics Canada.

The internationally defined weighting specification for PIRLS required that each assessed student's sampling weight be the product of six weighting factors: the inverse of the school's probability of selection, an adjustment for the school-level nonresponse, the inverse of the classrooms' probability of selection, an adjustment of classroom level nonresponse, the inverse of the student's probability of selection (always equal to 1 because whole classrooms were selected), and an adjustment for student-level nonresponse.

Because PIRLS used cluster sampling, conventional formulas for estimating sampling variability that assume simple random sampling and hence independence of observation are

inappropriate. For this reason, PIRLS uses a jackknife repeated replication method to estimate standards errors that capture sampling variance. The replicate weights to estimate correct standard errors are not explicitly included in the data files. Instead, users must use special macros developed to both produce and apply the replicate weights. These macros are available in the *PIRLS 2006 User Guide for the International Database*, chapters 4 (for SPSS) and 5 (for SAS). In addition, special software that works in conjunction with SPSS will generate these weights and apply them in analysis. This special software is the IDB Analyzer program, and can be downloaded from the TIMSS and PIRLS International Study Center website at http://timssandpirls.bc.edu/pirls2006/user_guide.html. Section 10.2 of chapter 10 describes IDB Analyzer further.

Chapter 9.
The PIRLS 2006 Data

9.1 PIRLS 2006 International Data

Data from PIRLS 2006 for all countries can be obtained from the PIRLS International Study Center at Boston College. The website can be accessed at http://timssandpirls.bc.edu/pirls2006/user_guide.html.

The International Database contains the following folders:

- Almanacs
- Codebooks
- Curriculum
- SAS Data
- SPSS Data
- IDB Analyzer
- Items
- Programs

Users should note that the IDB Analyzer is only available through the international website.

9.2 U.S. National Data Files

All of the files on the PIRLS NCES website contain data for the United States only. A CD-ROM containing the files, along with an Electronic Codebook, is also available from NCES. The file names follow the international naming conventions described in the *PIRLS 2006 User Guide for the International Database and Supplements* (edited by Pierre Foy and Ann M. Kennedy) which can be found at http://timssandpirls.bc.edu/pirls2006/user_guide.html.

The U.S. files available on the NCES website are the following:

- Student data
  - The data are contained in ASGUSAR2.DAT. This file contains plausible values for all scales from the reading assessment; questionnaire items and derived variables and index scores based on the student questionnaire; and student sampling weights and replicate weights. TOTWGT is the primary student-level weight and sums to the student population size in the United States.
  - There are 5,190 cases in this file (9 students completed the assessment but have no questionnaire data). Since the data are hierarchical (students are clustered with classrooms and schools), each student record contains identification variables that enable the user to merge the student, teacher, and school data.

  - An SPSS syntax file, ASGUSAR2.SPS.
  - A SAS syntax file, ASGUSAR2.SAS.
  - A codebook file (ASGUSAR2_CODEBOOK_PUD.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and unweighted frequencies.

- School data
  - The data are contained in ACGUSAR2.DAT. This file contains items from the school questionnaire, derived variables and index scores based on the school questionnaire, and the school weight, SCHWGT which is designed for school-level analysis. There are 183 cases in this file (1 school did not complete the questionnaire and therefore has missing data).
  - An SPSS syntax file, ACGUSAR2.SPS.
  - A SAS syntax file, ACGUSAR2.SAS.
  - A codebook file (ACGUSAR2_CODEBOOK_PUD.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and unweighted frequencies.

- Teacher data
  - The data are contained in ATGUSAR2.DAT. This file contains items from the teacher questionnaire and derived variables based on the teacher questionnaire. There are 253 cases in this file; 4 of those cases have missing data because the teachers did not complete the questionnaire. There is no weight variable on the teacher questionnaire data file because these variables are not designed to be analyzed independently. The teacher data can be merged with the student data using the student-teacher link file described below.
  - An SPSS syntax file, ATGUSAR2.SPS.
  - A SAS syntax file, ATGUSAR2.SAS.
  - A codebook file (ATGUSAR2_CODEBOOK_PUD.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and unweighted frequencies.

- Student-Teacher link file
  - The data are contained in ASTUSAR2.DAT. This file contains items needed to link students and teachers. There are 5,190 cases in this file. This file contains a special weight variable: the TCHWGT variable is specifically designed for using teacher background data in student-level analysis. The sole purpose of this file is to link teacher-level data with student-level data to perform appropriate student-level analysis.
  - An SPSS syntax file, ASTUSAR2.SPS.
  - A SAS syntax file, ASTUSAR2.SAS.

  - A codebook file (ASTUSAR2_CODEBOOK_PUD.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and unweighted frequencies.

- Cognitive (assessment) item data
  - The data are contained in ASAUSAR2.DAT. This file contains student responses to each item in the assessment. Note that some of the items have not been released, so there is little descriptive information about them. There are 5,190 cases in this file.
  - An SPSS syntax file, ASAUSAR2.SPS.
  - A SAS syntax file, ASAUSAR2.SAS.
  - A codebook file (ASAUSAR2_CODEBOOK_PUD.PDF) that includes variable names, variable location and format information, variable labels, question text, values, and unweighted frequencies.
- Questionnaires
  - The U.S. version of the student questionnaire is in the file P06_STQ_USA.PDF.
  - The U.S. version of the teacher questionnaire is in the file P06_TQ_USA.PDF.
  - The U.S. version of the school questionnaire is in the file P06_SQ_USA.PDF.
  - No U.S. version of the parent questionnaire was administered.
- Released items used in PIRLS 2006
  - Passages that were released after the 2006 administration of PIRLS can be found in P06_UG_SamplePassages1_Items.pdf.
- PIRLS_2006 data user's guide
  - This document, that you are reading now, PIRLS_2006_Data_Analysis_Users_Guide.pdf, contains information on the conduct of PIRLS in the United States. For complete, detailed information on using PIRLS data, users should consult the *PIRLS 2006 User Guide for the International Database and Supplements* (edited by Pierre Foy and Ann M. Kennedy) which can be found at http://timssandpirls.bc.edu/pirls2006/user_guide.html.

9.3 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" (variables used only in the United States). Note that the same assessment items were used by all countries. All variables collected internationally were included on U.S. instruments, except that the United States did not conduct a survey of parents. Variables used only in the United States are listed here:

Variables used only in the United States

Student file

RACE Race was collected from school records. Schools were first asked to indicate whether a student was Hispanic or non-Hispanic, and then whether they belonged to one or more of the following racial groups: White, Black, Asian, American Indian or Alaskan Native, Native Hawaiian or other Pacific Islander. The recoded categories in the data are: White (non-Hispanic), Black (non-Hispanic), Asian (non-Hispanic), American Indian/Alaskan Native (non-Hispanic) and Multi-racial or Other race (non-Hispanic).

School file

FRLUNC Percentage of students receiving free or reduced-price lunch based on sampling frame information.

9.4 Confidentiality

The PIRLS 2006 data are hierarchical and include school, teacher, 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 PIRLS International Study Center and NCES would not allow identification of individual U.S. schools, teachers, or students when compared against other public-use data collections. Disclosure limitations included identifying and masking potential disclosure risks to PIRLS schools and including an additional measure of uncertainty to school, teacher, and student identification through random swapping of data elements within the student and school files.

Chapter 10.
Using the PIRLS 2006 Data Files

10.1 Special Considerations—Plausible Values and Variance Estimates

Three aspects of the design of PIRLS 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 PIRLS is provided in the *PIRLS 2006 Technical Report.*

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, many statistical software packages operating on the assumption of a simple random sample will produce biased estimates of standard errors. Special procedures that use the sample design information and jackknife procedures are needed to correctly estimate standard errors.

The third aspect arises from the design of PIRLS and the use of plausible values for secondary analysis. In PIRLS, as in many national assessments, students are not administered every assessment item. Each item thus has missing student responses, although these are missing by design. For the purposes of secondary analysis, each student record has five "plausible values" for each score that is estimated; however, these are provided solely so that data users can calculate scale scores for student groups defined by the analyst. What this means, in effect, is that any secondary analysis 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, which cannot be simply averaged across the five plausible values.

10.2 The IDB Analyzer

The TIMSS and PIRLS International Study Center has developed easy-to-use software called the IDB Analyzer that is designed to deal with the three challenges listed above. The software can be downloaded without charge from the website at http://timssandpirls.bc.edu/pirls2006/user_guide.html.

The IDB Analyzer is designed to merge data across files (student, teacher, school) with ease, use the correct weight variables for each level of analysis, and produce variance estimates that take into account the plausible values and the sampling design of the study. The version available at the time of this posting requires that SPSS be installed on the computer as well as the IDB Analyzer (IDB Analyzer, though a separate program, uses SPSS routines in its programming). Users will also need to download the .dat and .sps files provided on the NCES website to read into the IDB Analyzer.

Directions for using the IDB Analyzer can be found in the *PIRLS 2006 User Guide for the International Database* which can be downloaded from the same website.

For users who do not wish to use the IDB Analyzer, the SPSS and SAS syntax files provided on the NCES website can be used to construct student, teacher, school, and assessment

data files. Information on constructing macros for use with SPSS or SAS can be found in the *PIRLS 2006 User Guide for the International Database*, chapters 4 (SPSS) and 5 (SAS).

10.3 Variance Estimation Variables

The student file (ASG), assessment file (ASA), and student-teacher link file (AST) each contain JKZONE and JKREP variables. JKZONE is the jackknife sampling zone, or stratum, to which the student's school is assigned. JKREP is the jackknife replicate, or primary sampling unit, to which the student's school is assigned. These variables are to be used in student-level analysis.

The school file (ACG) contains JKCZONE and JKCREP variables which can be used for school-level analysis. JKCZONE is the jackknife sampling zone, or stratum, to which the school is assigned. JKCREP is the jackknife replicate, or primary sampling unit, to which the school is assigned.

Complete information on the weights and variance estimation variables is available and can be found beginning on page 36 of the *PIRLS 2006 User Guide for the International Database*. Please note that the IDB Analyzer is programmed to use these variables correctly.

10.4 Merging School, Teacher, and Student Data

The IDB Analyzer is programmed to use the correct ID variables to merge files.

For those electing to use SPSS, SAS, or other software, the major IDs are as follows:

- IDSCHOOL is a four-digit identification code that uniquely identifies the participating schools within each country. The school codes are not unique across countries.
- IDCLASS is a six-digit identification code that uniquely identifies the sampled classrooms within a country. The variable IDCLASS has a hierarchical structure and is formed by concatenating the IDSCHOOL variable and a two-digit sequential number identifying the sampled classrooms within a school.
- IDSTUD is an eight-digit identification code that uniquely identifies each sampled student in a country. The variable IDSTUD also has a hierarchical structure and is formed by concatenating IDCLASS variable and a two-digit sequential number identifying all students within each classroom.
- IDTEACH is a six-digit identification code that uniquely identifies a teacher within a school. It has a hierarchical structure and is formed by the concatenation of IDSCHOOL and a two-digit sequential number within each school. This variable is on the AST and ATG files only.
- IDLINK uniquely identifies the class for which a teacher answered a questionnaire. The combination of linking variables IDTEACH and IDLINK uniquely identifies all teacher-class combinations in the database. This variable is on the AST and ATG files only.

Complete information on IDs and linkages can be found in the PIRLS 2006 Users Guide for the International Database.

Given the complexity of analyses without the IDB Analyzer, when conducting analyses without the IDB Analyzer it is essential to replicate officially published PIRLS findings before conducting any new analyses.

References

Foy, P., and Kennedy, A.M. (2008). *PIRLS 2006 User Guide for the International Database and Supplements.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M.O., Mullis, I.V.S., and Kennedy, A.M. (Eds.). (2007). *PIRLS 2006 Technical Report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., and Foy, P. (2007). *IEA's Progress in International Reading Literacy Study in Primary School in 40 Countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

TIMSS & PIRLS International Study Center (2005). *PIRLS 2006 School Coordinator Manual.* Chestnut Hill, MA: Boston College.

UNESCO (United Nations Educational, Scientific, and Cultural Organization). (1999). *Operational Manual for ISCED-1997 (International Standard Classification of Education).* Paris: UNESCO Institute for Statistics.

Appendix A
Comparing PIRLS with NAEP in Reading

Comparing PIRLS and NAEP in Reading

Maria Stephens

American Institutes for Research

A.1 Background

Both the Progress in International Reading Literacy Study (PIRLS) 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 PIRLS 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 PIRLS and to compare results from PIRLS with recent findings from NAEP reading assessments.

NAEP measures fourth-, eighth-, and twelfth-grade students' performance in reading, mathematics, and science with assessments designed specifically for national and state information needs. Alternatively, PIRLS enables the United States to benchmark the performance of fourth-graders in reading literacy to that of other countries. Both assessments are conducted regularly to allow the monitoring of student outcomes over time.[2]

Although PIRLS and NAEP both assess reading skills among fourth-grade students, each was designed to serve a different purpose 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 explains the similarities and differences in these assessments and identifies what each assessment contributes to the overall knowledge base on student performance. To do so, it uses information from the 2006 administration of PIRLS and from the 2007 results from NAEP for fourth- and eighth-grade reading (see Baer et al. 2007; Grigg, Donahue, and Dion 2007; Grigg, Lauko, and Brockway 2006; Lee, Grigg, and Dion 2007; and Lee, Grigg, and Donahue 2007).

A.2 Comparing Features of the Assessments

PIRLS and NAEP differ from one another on several key features, including purpose, population, precision of estimates, and content.

[1] PIRLS is conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). NAEP is sponsored by the National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. The United States also participates in the Trends in International Mathematics and Science Study (TIMSS), 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 the most recent TIMSS (TIMSS 2003) and NAEP.

[2] All statements about NAEP in this paper refer to national NAEP (versus long-term trend NAEP). NAEP currently assesses fourth- and eighth-grade reading and mathematics every two years, and twelfth-grade reading and mathematics, as well as science at all three grades, every four years. PIRLS is on a five-year cycle.

A.2.1 Purpose and Proximity to Curriculum

The goals of the assessments 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.

PIRLS is the U.S. source for internationally comparative information on the reading achievement of students in the fourth grade and on related contextual aspects such as reading curricula and classroom practices across countries. The PIRLS framework and specifications are developed in a collaborative process involving international reading experts and national research coordinators from each participating country, thus the consensus of the international research community and the interests of a wide range of nations.

A.2.2 Population

PIRLS provides benchmarks associated with a diverse group of countries.

NAEP is designed to produce estimates of student achievement in the United States while PIRLS is designed to provide comparable assessment results across a range of developed and developing countries. The sponsor of PIRLS is the International Association for Evaluation of Educational Achievement (IEA), which includes in its assessments a diverse group of countries and jurisdictions. Thirty-six countries and 9 jurisdictions within countries participated in PIRLS 2006 and are included individually in the country rankings of student performance (table A-1). The 36 countries include 16 countries from the Organization of Economic Cooperation and Development (OECD), representing the world's most industrialized nations. Regions of specific OECD countries are also included as separate jurisdictions: five Canadian provinces, England and Scotland (within the United Kingdom), and the Flemish and French communities of Belgium. This means that in PIRLS, in some cases, the United States is being compared not just with other countries but with jurisdictions within countries. Twenty additional countries outside of the OECD were also part of PIRLS 2006.

Table A-1. Country and region participation in PIRLS

Country group	Participating countries	
OECD jurisdictions	Austria	Italy
	Belgium, Flemish region Belgium, French region	Luxemburg
		Netherlands
	Canada, Alberta	New Zealand
	Canada, British Columbia	Norway
	Canada, Nova Scotia	Poland
	Canada, Ontario	Slovak Republic
	Canada, Quebec	Spain
	Denmark	Sweden
	France	United Kingdom, England
	Germany	United Kingdom, Scotland
	Hungary	United States
	Iceland	
Other countries	Bulgaria	Macedonia
	Chinese Taipei	Moldova
	Georgia	Morocco
	Hong Kong	Qatar
	Indonesia	Romania
	Iran	Russian Federation
	Israel	Singapore
	Kuwait	Slovenia
	Latvia	South Africa
	Lithuania	Trinidad and Tobago

SOURCE: U.S. Department of Education, National Center for Education Statistics, Progress in International Reading Literacy (PIRLS) 2006

The students in the two studies represent slightly different groups.

NAEP and PIRLS are sample-based assessments—meaning that each program administers the assessment to a representative set of U.S. students. The set of students is selected in such a way that the results can be generalized to the larger population; however there are small differences between the two studies in sample selection procedures. In addition, although both NAEP and PIRLS selected samples of fourth-grade students, their populations differed slightly because NAEP selected a sample of 2007 fourth-graders and PIRLS selected a sample of 2006 fourth-graders.

A.2.3 Precision of Estimates

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

NAEP and PIRLS are all 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 PIRLS differ, however, in the size of the differences in performance they are intended to detect. Student performance varies widely across countries and so PIRLS is designed to detect only relatively large differences. NAEP is designed to detect smaller differences. NAEP can detect smaller variations in student performance within

the U.S. than PIRLS, 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 for both studies are calculated to balance needs for precision of estimates against burden to respondents. Because of NAEP's need for greater precision, NAEP samples many more students than does PIRLS (table A-2).

Table A-2. Sample sizes in NAEP and PIRLS

Study	Number of students sampled	Number of schools sampled
NAEP 2007 (4th grade)	191,000	7,830
PIRLS 2006	5,190	183

SOURCE: U.S. Department of Education, National Center for Education Statistics, Progress in International Reading Literacy (PIRLS) 2006 and National Assessment of Educational Progress (NAEP) 2007.

A.2.4 Content

The reading skills 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 Fourth-grade reading: PIRLS 2006 and NAEP 2007

To date, there have been two studies undertaken to compare NAEP and PIRLS in these two ways. The first study compared NAEP 2002 and PIRLS 2001 at both the framework and item levels and was documented in an NCES technical report (Binkley and Kelly 2003).[3] The second study updates the first with analysis of the passages and item sets added in NAEP 2007 and PIRLS 2006.[4] These studies form the basis of the text that follows.

Definitions and organization. In terms of how the domain is defined, there is considerable overlap between the NAEP and PIRLS concepts of reading literacy. For example, the frameworks for both assessments: (1) identify reading as a constructive process that involves interaction between the reader and the text; (2) suggest that the context for reading is an important element in how readers make meaning of text and in the skills and strategies they select; and (3) note that the structural elements of text influence readers' strategies. The

[3] The Binkley and Kelly (2003) working paper provides information on the theoretical comparisons of the fourth-grade reading definitions and frameworks; this is valid because the frameworks for each study have remained essentially the same.

[4] This study involved a five-person expert panel to externally verify the continuity of the NAEP 2002 and 2007 frameworks and to classify the new PIRLS items to the NAEP framework. This occurred in September 2007. The classifications for the new PIRLS items were combined with the classifications from the first study for the items used in both 2001 and 2006 to obtain a complete data set that could be compared with the NAEP assessment specifications. This informs the section on the item-level comparisons.

differences are relatively minor: the PIRLS framework is more explicit about its targeting to young readers and acknowledges a more diverse set of reading contexts such as for personal enjoyment (versus the NAEP framework, which focuses more on school-based reading and is intended to be generally applicable across younger to older grades).

In terms of the organization of the frameworks, both NAEP and PIRLS are organized around two-dimensional matrices, which specify *processes* (i.e., the cognitive element) and the *purposes* or contexts for which students read. In particular, there are some notable differences at the framework level in how the processes (called *aspects* in NAEP) are broken out and elaborated. NAEP's four categories include: forming a general understanding, developing an interpretation, making reader-text connections, and examining content and structure. PIRLS' four categories include: locating and retrieving explicitly stated information, making straightforward inferences, interpreting and integrating ideas and information, and examining and evaluating content, language and textual elements. The key areas of difference are that there is no apparent counterpart in the NAEP framework to the PIRLS locate and retrieve category, and there is no explicit counterpart in the PIRLS framework to the NAEP category that requires readers to think beyond the text and apply it to the real world (i.e., make reader-text connections). This suggests that there may be certain NAEP and PIRLS items that are unique to the respective programs.

In terms of the purposes for which students read, both frameworks specify a literary purpose and an information-related purpose. While the literary purposes seem to be defined in a similar way across the assessments, the information-related purposes suggest slight differences. PIRLS assesses not just reading to *acquire* information, but also to *use* information, in a way that goes beyond NAEP's definition. At the older grades, the NAEP framework includes a "reading to perform a task" purpose, which focuses on reading to learn how to do something, which is more similar to the use information aspect of PIRLS' "reading to acquire and use information" purpose.

Passage and item analyses. The types of passages included in NAEP and PIRLS reflect the purposes that are assessed. In NAEP, students are presented with short stories, legends, biographies, and folktales, as well as magazine articles that focus on people, places, and events of interest to children—to cover both its literary experience and information purposes. Similarly, PIRLS also presents narrative fiction, usually in the form of short stories, as well as informational articles and, distinct from NAEP, brochures to cover its two similar purposes. Both NAEP and PIRLS strive to be "authentic" in that they try to present passages and items that would be encountered in and out of school. NAEP specifically calls for the use of authentic texts, and all passages are shown as previously published and generally are not edited at all (in terms of content or formatting) for use in NAEP. PIRLS also strives to use previously published texts, but has a more liberal policy on editing and changing the format of the texts used—which is sometimes necessary in an international context in order to meet constraints of translation to multiple languages and for culturally diverse participants. U.S. experts who have examined the PIRLS passages have noted the more edited, and sometimes less continuous, nature of some of these than the NAEP passages, particularly among passages for information purpose.

Altogether, the NAEP and PIRLS fourth-grade assessments each include 10 reading passages, although each student receives only a subset of those passages.[5] In terms of length, the PIRLS passages tend to be shorter than the NAEP passages, averaging 707 words per passage compared to NAEP's 823 words per passage. The PIRLS passages range from 403 to 855 words; NAEP passages range from 644 to 1,361 words.

Readability analyses also suggest that the PIRLS passages may be slightly easier than NAEP (table A-3). On a very simple measure, for example, sentence counts show that the PIRLS passages, with a higher number of sentences per 100 word sample, consist of shorter sentences on average than do the NAEP passages. On other more elaborate measures, such as Fry and Flesch analyses, which use sentence count along with syllable count to determine a corresponding age and grade level for each text, PIRLS passages are calculated to be about one grade level below the NAEP passages. Finally, a Lexile measure, which indicates the reading demand of the text in terms of semantic difficulty (vocabulary) and syntactic complexity (sentence length) and which is more recently developed and normed than the other measures, also suggests that the PIRLS passages are suitable for one to two grades below those from NAEP. It should be noted, however, that both assessments do include a range of passages below and above the targeted grade level to capture a range of reading ability.

Table A-3. Results of readability analyses of PIRLS 2006 and NAEP 2007 passages

Characteristic	PIRLS	NAEP
Number of sentences per 100 words	8.16	7.16
Number of syllables per 100 words	132.0	132.6
Fry average age	10.30	11.07
Fry average grade level	5th	6th
Flesch average reading ease	Easy (82.3)	Fairly easy (79.7)
Flesch average grade level	5th–6th	7th
Lexile score	819.0	936.7
Lexile score corresponding grade level	4th–5th	6th–7th

SOURCE: U.S. Department of Education, Institute of Education Science, National Center for Education Statistics, Progress in International Reading Literacy Study (PIRLS) 2006, and National Assessment of Educational Progress (NAEP) 2007.

Each of these passages has a number of items associated with it—approximately 12 to 13 items per passage in PIRLS and 10 per passage in NAEP. Mapping the PIRLS items onto NAEP's cognitive processes, or aspects, and comparing these classifications with those for the NAEP items confirms some of the similarities and differences suggested by the forgoing framework analysis. The two assessments are similar in that the majority of items on both assessments require students to develop an interpretation about what they have read (rather than to recall information or other simpler reading tasks), although there is a greater emphasis on this in NAEP, with 69 percent of items classified as requiring interpretation compared to 60 percent of the PIRLS items. PIRLS also has a notably smaller percentage of items classified as forming a general understanding or making reader-text connections, having half or less the percentage NAEP has in those categories. One of the major differences between the two assessments, however, is that there are a number of PIRLS items (21 percent) that do not fit onto the NAEP

[5] This quantitative information in this paragraph and the two that follow was calculated as part of the passage and item analyses undertaken for this section.

framework at all. In nearly all cases, these are items that ask the reader to retrieve explicitly stated information, which is not a skill delineated in the NAEP framework or found in its items.

A.3 Results in the Context of Assessment Differences

Both PIRLS and NAEP provide a measure of fourth-grade reading. It is natural to compare their results, but the distinctions described previously need to be kept in mind in understanding the converging or diverging results.

A.3.1 Comparing Select Results for Fourth-Grade Reading

The most recent results from PIRLS and NAEP include information on trends over time in fourth-grade reading: in PIRLS between 2006 and 2001 and in NAEP between 2007 and several earlier time points going back to 1992. Here we describe the NAEP 2002 to 2007 period, since it provides a similar time interval to PIRLS.

PIRLS shows that statistically there is no change in U.S. fourth-grade students' average scores from 2001 to 2006. This contrasts with NAEP results for 2007, which show an upward tick (by 2 score points) in fourth-grade reading scores from 2002, all of which occurred since 2005. However, although the populations in PIRLS and NAEP are the same, as the previous sections highlighted, there are some differences in the nature of the reading passages and in the reading skills being measured, with about one-fifth of the PIRLS items not corresponding well to the NAEP framework. Additionally, because NAEP uses a much larger sample size, it is more sensitive to picking up small changes over short periods of time than is PIRLS, which is not designed primarily for that purpose but for detecting differences among countries.

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 PIRLS add value by providing a method for comparing U.S. performance to the performance of students in other nations. However, differences between studies need to be recognized when interpreting results. Some of the differences between NAEP and PIRLS 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 PIRLS, the content of which is determined internationally in collaboration with other countries and reflecting consensus views of key content.
- PIRLS provides benchmarks with different groups of countries. Thirty-six countries participated in PIRLS, 16 of which are industrialized OECD countries and 20 of which were from outside of the OECD
- The students being studied represent different groups. Both NAEP and PIRLS use grade-based samples and both target fourth-grade students. However, the last NAEP

assessment in fourth-grade reading was in 2007, whereas for PIRLS it was in 2006, so the results do not generalize to the same group of students.

- The assessments are designed to measure student performance at different levels of precision. NAEP and PIRLS 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 measures performance at a higher level of precision than PIRLS. These differences can have an impact on the assessments' sensitivities in detecting changes in student performance.
- The reading skills 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 reading frameworks of NAEP and PIRLS show areas of potential overlap and potential difference in terms of the content and skills being measured in the respective subject areas and grades. Further, additional analyses of the fourth-grade reading passages and items show that (1) PIRLS passages are slightly shorter and slightly easier than NAEP fourth-grade passages, and (2) PIRLS appears to have a subset of items that are distinct from the types of items found in NAEP.

A.5 References

Baer, J., Baldi, S., Ayotte, K., Green, P., and McGrath, D. (2007). *The Reading Literacy of U.S. Fourth-Grade Students in an International Context: Results from the 2001 and 2006 Progress in Reading Literacy Study (PIRLS)* (NCES 2008-017). Washington, DC: National Center for Education Statistics, Institute of Education Science, U.S. Department of Education.

Binkley, M., and Kelly, D.L. (2003). *A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments* (Working Paper No. 2003:10). Washington, DC: National Center for Education Statistics, Institute of Education Science, U.S. Department of Education.

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 that is 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 nonresponse bias. One way to characterize and quantify the presence of 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. The National Center for Education Statistics (NCES) requires a nonresponse bias analysis for all datasets based on surveys in which the unit response rate is less than 85 percent. This report presents the results of a nonresponse bias analysis for the 2006 PIRLS (Progress in International Reading Literacy Study). NCES also requires an item nonresponse bias analysis for all questionnaire items for which the response rate is less than 85 percent. For the 2006 PIRLS all items had a response rate of 85 percent or higher, so an item nonresponse analysis is not required.

The objective of this nonresponse bias analysis report 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. Furthermore, the response rate in this case was below the NCES 85 percent threshold, thus posing a situation in which nonresponse bias could be an important quality component of the final data.

B.1.2 Overview of PIRLS

PIRLS is a continuing assessment of the reading comprehension of students in their fourth year of schooling in countries around the world. PIRLS not only helps responding countries understand the literacy skills of their students but also places the literacy of young readers within an international context. Drawing comparisons between countries reveals areas of strength and areas in need of improvement, offering countries insight into how the reading literacy of their students may be enhanced. PIRLS is conducted by the International Association for the Evaluation of Educational Achievement, with national sponsors in each responding country. In the United States, PIRLS is sponsored by NCES, in the Institute of Education Sciences in the U.S. Department of Education.

PIRLS 2006 was the second cycle of the study, which was first administered in 2001. PIRLS 2006 included 45 countries and subnational entities with well-defined education systems separate from their national systems (such as Canadian provinces and the Flemish and French-speaking communities of Belgium).

B.1.3 Sampling

Within each responding country, a minimum of 150 schools were selected for PIRLS 2006. In the United States, the PIRLS sample was designed to be representative of all fourth-grade students in the 50 states and the District of Columbia. In addition to the base sample (designed to yield 150 responding schools), the United States sampled additional private schools and high-poverty schools (defined as those schools in which 50 percent or more of students are eligible to receive free or reduced-price lunch). The U.S. sample was designed to yield 180 responding schools. This oversampling was designed to provide sufficient information on these two groups of schools, which are of special interest to researchers and policymakers.

The U.S. school sample for PIRLS 2006 was drawn in March 2005. The sampling frame was constructed using data from the 2002–03 Common Core of Data and preliminary data from the 2003–04 Private School Universe Survey. To be consistent with the sampling design for PIRLS 2001, the frame was divided into two explicit strata as follows: (1) a Metropolitan Statistical Area (MSA) stratum was created that included schools located in the 10 most populous MSAs according to the U.S. Census Bureau; and (2) all schools outside those MSAs were grouped into 451 Primary Sampling Units (PSUs)[1] by sorting on MSA and then by county. PSUs were designed to fit within state boundaries and, where possible, within county and city boundaries. Before the selection process, schools were sorted by state, percentage of racial/ethnic minority students, control of school (public/private), percentage of students eligible for free or reduced-priced lunch, and locale. Within each PSU and MSA stratum, schools were selected on the basis of the number of fourth-grade students in the school, so that larger schools had a higher probability of selection than smaller schools. The final sample included 214 schools; 150 were chosen from PSUs and 64 were selected from the MSA stratum.

B.1.4 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 replaced. 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 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

[1] PSUs are the first-level sampling units. They are geographic areas consisting of counties or groups of counties.

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 allow for comparisons of these dependent samples. In these tables, the bias is computed as 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.

Section B.2 implements this analysis for the original sample of 214 schools attempting to identify any bias in the group of 120 responding schools. Section B.3 repeats the analysis for the same-sized sample of 214 schools but this time there are 183 responding schools, of which 63 are substitute schools. In each section the analysis first studies categorical variables using a chi-square statistic to test differences between responding and nonresponding schools. The second part of the analysis focuses on continuous variables and uses the *t* statistic to test differences in means between respondents and nonrespondents. The analysis uses the comparison between respondents and nonrespondents to uphold the assumptions of the *t* tests and chi-square tests which require two independent samples. As an additional effort to identify any presence of bias, we employ logistic regression analysis with response status as the dependent variable.

B.2 Nonresponse Bias Analysis: Original Sample

This section presents the results of the nonresponse bias analysis, based on the original Progress in International Reading Literacy Study (PIRLS) 2006 sample of 214 eligible U.S. schools consisting of 120 responding schools and 94 nonresponding schools. Typically, in most National Center for Education Statistics 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.4 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 first analysis compares the distribution of the 120 responding schools with that of the 94 nonresponding schools using weighted data in each case for four 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 120 responding schools with the 214 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.4 (under Weighting, B.4.2.3).

B.2.1 Categorical Variables—Before Nonresponse Weight Adjustment

The categorical variables used in this set of analyses include four school characteristics: school control, community type, Census region, and poverty level. The distributions of nonresponding and responding schools in the U.S. PIRLS 2006 original sample are shown in table B-1 by the four school characteristics *before* the nonresponse weight adjustment. The hypothesis of independence between the school characteristics and response status was tested using chi-square statistics at the 5 percent level. The *p* values for the chi-square tests are presented in the table. The relative bias is also presented in the table and is calculated as the bias divided by the estimate from the responding sample.

Based on these analyses, no measurable differences were detected between the responding and nonresponding schools along the four school characteristics cited above. However, 6 of the 11 computed relative biases are greater than 10 percent, including 3 of the 4 categories of Census region. This indicates potential bias even though no statistically significant differences were detected.

B.2.2 Categorical Variables—After Nonresponse Weight Adjustment

The distributions of the responding and full sample schools in the U.S. PIRLS 2006 original sample are shown in table B-2 by the four school characteristics *after* the nonresponse weight adjustment. The hypothesis of independence between the school characteristics and response status was tested using chi-square statistics at the 5 percent level. The *p* values for the chi-square tests and the relative bias are presented in the table.

Based on these analyses, no measurable differences were detected between the responding and full sample schools along the four school characteristics considered: school control, community type, Census region, and poverty level. Compared with bias before weight adjustment, the absolute value of the bias after weight adjustment decreased for all school characteristics, except for the central city community type and the Central Census region. Also, only 2 of the 11 computed relative biases after weight adjustment are greater than 10 percent.[2]

B.2.3 Continuous Variables—Before Nonresponse Weight Adjustment

The continuous variables used in this set of analyses include student enrollment, student race/ethnicity, and the percentage of students eligible for free or reduced-price lunch. It is important to note that student enrollment is at the school level, and student race/ethnicity and percentage of students eligible for free or reduced-price lunch are averaged over all schools for the respondents and nonrespondents. Thus the race/ethnicity variables are the proportion of each ethnic group across schools. Similarly, the free or reduced-price lunch variable is the proportion of students who have access to such a program across schools. Because it is continuous, results for the free or reduced-price lunch variable differ from those for the categorical poverty variable

[2] The 10 percent cutoff is subjective and was chosen to describe what appear to be large relative biases. The relative biases are shown in the tables, so the reader can choose different criteria to help determine the effect of the bias on the data.

(presented in table B-1). Means for each continuous variable for nonresponding and responding schools in the original sample before the nonresponse weight adjustment are shown in tables B-3 through B-5. The difference between the means of responding and nonresponding schools was tested using a *t* test. The *p* values for the *t* tests and the relative bias are presented in the tables.

Table B-1. Percentage distribution of nonresponding and responding schools in the U.S. PIRLS original sample before nonresponse weight adjustment, by selected school characteristics: 2006

	Sample schools				
	Responding (standard error)	Nonresponding (standard error)	Bias	Relative bias	Chi–square *p* value
School control					0.7360
Public	81.97	84.39			
	(4.86)	(5.32)	2.42	0.0295	
Private	18.03	15.61			
	(4.86)	(5.32)	–2.42	–0.1342	
Community type					0.8755
Central city	32.04	32.68			
	(4.68)	(5.96)	0.64	0.0110	
Urban fringe/large town	40.75	44.38			
	(5.29)	(6.48)	3.63	0.0891	
Rural/small town	27.20	22.94			
	(5.23)	(7.21)	–4.26	–0.1566	
Census region					0.4724
Northeast	18.59	26.05			
	(2.86)	(6.03)	7.46	0.4013	
Southeast	38.56	28.69			
	(4.07)	(4.71)	–9.87	–0.2560	
Central	20.90	24.44			
	(2.80)	(5.38)	3.54	0.1694	
West	21.94	20.83			
	(4.02)	(4.45)	–1.11	–0.0506	
Poverty level					0.3458
High	30.79	24.52			
	(4.46)	(4.76)	–6.27	–0.2037	
Low	69.21	75.48			
	(4.46)	(4.76)	6.27	0.0906	

NOTE: Detail may not sum to totals because of rounding. For public schools, a high-poverty school is defined as one in which 50 percent or more of the students are eligible for response in the National School Lunch Program; all private schools are treated as low-poverty schools. Census region is the state-based region of the country (see section B.4 for a listing of states by Census region). 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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-2. Percentage distribution of full sample and responding schools in the U.S. PIRLS original sample after nonresponse weight adjustment, by selected school characteristics: 2006

	Sample schools				
	Responding (standard error)	Full sample (standard error)	Bias	Relative bias	Chi–square *p* value
School control					0.7343
Public	81.95 (4.84)	82.96 (3.61)	1.01	0.0123	
Private	18.05 (4.84)	17.04 (3.61)	–1.01	–0.0560	
Community type					0.9711
Central city	31.54 (4.68)	32.30 (3.67)	0.76	0.0241	
Urban fringe/large town	42.74 (5.46)	42.24 (4.08)	–0.50	–0.0117	
Rural/small town	25.72 (5.06)	25.46 (4.25)	–0.26	–0.0101	
Census region					0.6141
Northeast	19.05 (2.36)	21.64 (2.52)	2.59	0.1360	
Southeast	37.74 (3.89)	34.53 (2.81)	–3.21	–0.0851	
Central	30.67 (2.16)	22.35 (2.15)	–8.32	–0.2713	
West	22.54 (4.02)	21.49 (2.70)	–1.05	–0.0466	
Poverty level					0.4922
High	30.19 (4.48)	28.23 (3.28)	–1.96	–0.0650	
Low	69.81 (4.48)	71.77 (3.28)	1.96	0.0281	

NOTE: Detail may not sum to totals because of rounding. For public schools, a high-poverty school is defined as one in which 50 percent or more of the students are eligible for response in the National School Lunch Program; all private schools are treated as low-poverty schools. Census region is the state-based region of the country (see section B.4 for a listing of states by Census region). 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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-3. Mean enrollment of nonresponding and responding schools in the U.S. PIRLS original sample before nonresponse weight adjustment: 2006

Student enrollment	Sample schools: Responding (standard error)	Sample schools: Nonresponding (standard error)	Bias	Relative bias	*t* test *p* value
Total school	427.56	456.81			
	(21.38)	(36.48)	29.25	0.0684	0.0361
Fourth grade	66.50	76.96			
	(4.00)	(6.78)	10.46	0.1573	0.0238

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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-4. Mean percentage of students in nonresponding and responding schools in the U.S. PIRLS original sample before nonresponse weight adjustment, by race/ethnicity: 2006

Race/ethnicity	Sample schools: Responding (standard error)	Sample schools: Nonresponding (standard error)	Bias	Relative bias	*t* test *p* value
Asian or Pacific Islander	2.39	4.78			
	(0.34)	(0.92)	2.39	1.0000	0.0174
Black, non-Hispanic	14.54	15.39			
	(2.09)	(1.95)	0.85	0.0585	0.7692
Hispanic	14.43	21.91			
	(1.97)	(2.90)	7.48	0.5184	0.0433
American Indian or Alaska Native	2.85	0.93			
	(1.32)	(0.32)	–1.92	–0.6737	0.1594
White, non-Hispanic	53.49	53.45			
	(3.12)	(3.27)	–0.04	–0.0000	0.9938
Other	12.30	3.53			
	(2.39)	(1.07)	–8.77	–0.7130	0.0011

NOTE: Data on race/ethnicity were missing for 11 public schools. 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. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-5. Mean percentage of public school students eligible for free or reduced-price lunch in nonresponding and responding public schools in the U.S. PIRLS original sample before nonresponse weight adjustment: 2006

Students	Sample schools		Bias	Relative bias	*t* test *p* value
	Responding (standard error)	Nonresponding (standard error)			
Percentage of students eligible for free or reduced-price lunch	34.13 (2.82)	36.24 (3.05)	2.11	0.0618	0.6163

NOTE: Free or reduced-price lunch information was available only for public schools. 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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Significant differences were limited to total and fourth-grade enrollment and three racial/ethnic categories. Responding schools had fewer total and fourth-grade students than nonresponding schools. There were higher percentages of Hispanic and Asian or Pacific Islander students in the nonresponding schools than in responding schools and a higher percentage of Other students in responding than in nonresponding schools. However, the relative bias was about 67 percent for percentage of American Indian students, which indicates potential bias even though no statistically significant differences were detected.

B.2.4 Continuous Variables—After Nonresponse Weight Adjustment

Means for each continuous variable for the responding and full sample schools in the original sample after the nonresponse weight adjustment are shown in tables B-6 through B-8. The difference between the means of responding schools and the full sample was tested using a *t* test. The *p* values for the *t* tests and the relative bias are presented in the tables.

Significant differences were limited to two racial/ethnic categories. There were higher percentages of Asian or Pacific Islander students in the full sample than in responding schools and a higher percentage of Other students in responding schools than in the full sample. Compared with bias before weight adjustment, the absolute value of the bias after weight adjustment decreased for all school characteristics, except for percentage of White students. Also, the relative biases after weight adjustment is about 19 percent for percentage of Hispanic students and about 20 percent for percentage of American Indian students.

Table B-6. Mean enrollment of full sample and responding schools in the U.S. PIRLS original sample after nonresponse weight adjustment: 2006

Student enrollment	Sample schools: Responding (standard error)	Sample schools: Full sample (standard error)	Bias	Relative bias	*t test* *p* value
Total school	432.69	439.52			
	(21.70)	(19.27)	6.83	0.0158	0.6919
Fourth grade	67.24	70.77			
	(4.03)	(3.58)	3.53	0.0525	0.2530

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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-7. Mean percentage of students in full sample and responding schools in the U.S. PIRLS original sample, by race/ethnicity after nonresponse weight adjustment: 2006

Race/ethnicity	Sample schools: Responding (standard error)	Sample schools: Full sample (standard error)	Bias	Relative bias	*t test* *p* value
Asian or Pacific Islander	2.48	3.41			
	(0.35)	(0.44)	0.93	0.3750	0.0360
Black, non-Hispanic	14.33	14.91			
	(2.05)	(1.45)	0.58	0.0405	0.6427
Hispanic	14.84	17.62			
	(1.92)	(1.61)	2.78	0.1873	0.0875
American Indian or Alaska Native	2.56	2.03			
	(1.13)	(0.77)	–0.53	–0.2070	0.2057
White, non-Hispanic	54.22	53.47			
	(3.09)	(2.23)	–0.75	–0.0138	0.7137
Other	11.58	8.57			
	(2.20)	(1.46)	–3.01	0.2600	0.0057

NOTE: Data on race/ethnicity were missing for 11 public schools. 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. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-8. Mean percentage of public school students eligible for free or reduced-price lunch in full sample and responding public schools in the U.S. PIRLS original sample after nonresponse weight adjustment: 2006

Students	Sample schools		Bias	Relative bias	t test *p* value
	Responding (standard error)	Full sample (standard error)			
Percentage of students eligible for free or reduced-price lunch	33.91 (2.84)	35.03 (2.09)	1.12	0.0330	0.5456

NOTE: Free or reduced-price lunch information was available only for public schools. 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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

B.2.5 Logistic Regression Model

To examine the joint relationship of various characteristics to school nonresponse, the analysis employed a logistic regression model with response status as the binary dependent variable and frame characteristics as predictor variables. The null hypothesis is that there is no difference between responding and nonresponding schools.

The equation was first modeled in SAS using the stepwise procedure with the design weight[3] and an alpha = 0.05 inclusion criteria. That is, predictor variables were added one at a time, and if a predictor variable was significant at the 5 percent level, then it was kept in the model. Public and private schools were modeled together using the variables available for all schools; private schools were considered low-poverty schools. All variables included in this initial model were significant and 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. The school characteristics significantly different between responding and nonresponding schools were the percentage of Asian or Pacific Islander students, which was higher in nonresponding schools than in responding schools, schools that had fewer than 62 fourth-graders, and schools containing between 62 and 85 fourth-graders. These latter two parameters had higher estimates for responding schools than for nonresponding schools, which is inconsistent with the results in table B-3 due to the fourth-grade enrollment variable being defined differently. The SUDAAN estimates, standard errors, test statistics, and *p* values are reported in table B-9 for the intercept and significant variable.

[3] This analysis was only done using the design weight for respondents and nonrespondents before weight adjustment and not using the final weight for respondents after weight adjustment because the model uses response status as the dependent variable.

Table B-9. Logistic regression reduced model parameters using the U.S. PIRLS original sample: 2006

Parameter	Estimate	Standard error	Test for H_0: parameter = 0	*p* value
Intercept	0.21	1.63	0.13	0.8996
Percent Asian or Pacific Islander students	-0.09	0.03	-2.81	0.0054
Category 1: Schools that had less than 62 fourth-graders	1.72	0.80	2.16	0.0317
Category 2: Schools containing between 62 and 85 fourth-graders	1.44	0.64	2.27	0.0244
Category 3: Schools containing between 85 and 129 fourth-graders	0.28	0.50	0.57	0.5721
Census Region: Northeast	-0.82	0.65	-1.25	0.2126
Census Region: Southeast	0.04	0.51	0.07	0.9454
Census Region: Central	-0.54	0.97	-0.80	0.4261
Percent of Eligible Students receiving Free or Reduced Lunch	-0.01	0.01	-0.98	0.3266
High Poverty Status	1.18	0.76	1.57	0.1184
Central city	0.38	0.63	0.60	0.5472
Urban fringe/large town	0.42	0.60	0.70	0.4816
Total Students Enrolled in School	0.00	0.00	1.13	0.2588
Publicly Controlled School	0.41	0.73	0.56	0.5765
Percent American Indian Students	0.01	0.02	0.47	0.6356
Percent White Students	-0.02	0.01	-1.63	0.1042
Percent Hispanic Students	-0.02	0.01	-1.63	0.0656
Percent Black Students	-0.02	0.01	-1.63	0.0957

NOTE: H_0 = null hypothesis.
SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

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

To reach acceptable levels of response rate, following Progress in International Reading Literacy Study (PIRLS) guidelines, all sampled schools were linked with two potential substitute schools similar in stratification characteristics. As part of the sample design and field procedures guidelines set by PIRLS, a certain number of these substitute schools could be used to substitute for nonresponding schools. (See the technical notes in section B.4. for further information about weighting and the use of substitute schools.)

This section presents the nonresponse bias analysis based on the final sample of 214 schools selected for PIRLS 2006 in the United States. The total sample size is the same as the sample analyzed in section B.2 but the composition of responding and nonresponding schools differs. This time, in addition to the 120 responding original schools, there are 63 responding substitute schools, resulting in 183 responding schools and 31 nonresponding schools. The first analysis compares the distribution of the 183 responding sample schools, including 63 responding substitute schools, with the distribution of the 31 nonresponding schools using weighted data in each case. The weights are based on the design weight, which is the inverse of the probability of selection of the school. Eligible original schools that refused to participate in the survey and were not successfully replaced by a substitute are treated as nonrespondents. All

other eligible original sample schools were treated as respondents. The second analysis compares the distribution of the 183 responding schools with the 214 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.4 (under Weighting, B.4.2.3).

B.3.1 Categorical Variables—Before Nonresponse Weight Adjustment

The categorical variables used in this set of analyses include the same four school characteristics that were used in the analyses with the original respondent sample: school control, community type, Census region, and poverty level. The distribution of nonresponding and responding schools by the four school characteristics before nonresponse weight adjustment is shown in table B-10. The hypothesis of independence between the school characteristics and response status was tested using chi-square statistics at the 5 percent level. The *p* values for the chi-square tests are presented in the table. The relative bias is also presented in the table and is calculated as the bias divided by the estimate from the responding sample.

Based on these analyses, school control was found to be statistically significant. That is, public schools tended to be underrepresented and private schools overrepresented among the responding schools compared with the nonresponding schools. However, six of the nine computed relative biases for the other three school characteristics (community type, Census region, and poverty level) are greater than 10 percent. This indicates potential bias even though no statistically significant differences were detected.

B.3.2 Categorical Variables—After Nonresponse Weight Adjustment

The distribution of the responding and full sample schools by the four school characteristics after nonresponse weight adjustment is shown in table B-11. The hypothesis of independence between the school characteristics and response status was tested using chi-square statistics at the 5 percent level. The *p* values for the chi-square tests and the relative bias are presented in the table.

Based on these analyses, no measurable differences were detected between the responding schools and the full sample along the four school characteristics considered: school control, community type, Census region, and poverty level. Compared with bias before weight adjustment, the absolute value of the bias after weight adjustment decreased for all school characteristics, except for the Southeast Census region. Also, only 1 of the 11 computed relative biases after weight adjustment is greater than 10 percent.

Table B-10. Percentage distribution of nonresponding and responding schools in the U.S. PIRLS final sample before nonresponse weight adjustment, by selected school characteristics: 2006

School characteristic	Sample schools		Bias	Relative bias	Chi-square *p* value
	Responding (standard error)	Nonresponding (standard error)			
School control					0.0333
Public	81.41	95.18			
	(3.99)	(4.69)	13.77	0.1691	
Private	18.59	4.82			
	(3.99)	(4.69)	–13.77	–0.7407	
Community type					0.1351
Central city	30.71	38.92			
	(3.93)	(10.35)	8.21	0.2673	
Urban fringe/large town	41.03	49.50			
	(4.38)	(10.23)	8.47	0.2064	
Rural/small town	28.26	11.58			
	(4.69)	(6.10)	–16.68	–0.5902	
Census region					0.5687
Northeast	20.60	30.51			
	(2.96)	(9.92)	9.91	0.4811	
Southeast	34.45	34.47			
	(3.20)	(9.01)	0.02	0.0000	
Central	23.19	15.29			
	(2.64)	(7.14)	–7.90	–0.3401	
West	21.75	19.74			
	(3.11)	(7.04)	–2.01	–0.0924	
Poverty level					0.4525
High	25.09	32.38			
	(3.39)	(9.12)	7.29	0.2906	
Low	74.91	67.62			
	(3.39)	(9.12)	–7.29	–0.0973	

NOTE: Detail may not sum to totals because of rounding. For public schools, a high-poverty school is defined as one in which 50 percent or more of the students are eligible for response in the National School Lunch Program; all private schools are treated as low-poverty schools. Census region is the state-based region of the country (see section B.4 for a listing of states by Census region). 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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-11. Percentage distribution of full sample and responding schools in the U.S. PIRLS final sample after nonresponse weight adjustment, by selected school characteristics: 2006

School characteristic	Sample schools: Responding (standard error)	Sample schools: Full sample (standard error)	Bias	Relative bias	Chi-square *p* value
School control					0.4864
Public	80.58	82.97			
	(4.11)	(3.61)	2.39	0.0297	
Private	19.42	17.03			
	(4.11)	(3.61)	–2.39	–0.1231	
Community type					0.9517
Central city	31.17	31.64			
	(3.99)	(3.70)	0.47	0.0151	
Urban fringe/large town	42.71	41.99			
	(4.46)	(4.06)	–0.72	–0.0169	
Rural/small town	26.12	26.38			
	(4.41)	(4.26)	0.26	0.0100	
Census region					0.9777
Northeast	20.71	21.72			
	(2.52)	(2.52)	1.01	0.0488	
Southeast	34.20	34.46			
	(3.06)	(2.83)	0.26	0.0076	
Central	23.03	22.29			
	(2.31)	(2.15)	–0.74	–0.0321	
West	22.06	21.53			
	(3.09)	(2.70)	–0.53	–0.0240	
Poverty level					0.1889
High	24.64	25.91			
	(3.36)	(3.18)	1.27	0.0515	
Low	75.36	74.09			
	(3.36)	(3.18)	–1.27	–0.0169	

NOTE: Detail may not sum to totals because of rounding. For public schools, a high-poverty school is defined as one in which 50 percent or more of the students are eligible for response in the National School Lunch Program; all private schools are treated as low-poverty schools. Census region is the state-based region of the country (see section B.4 for a listing of states by Census region). 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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

B.3.3 Continuous Variables—Before Nonresponse Weight Adjustment

The continuous variables used in this set of analyses are the same ones that were used in the analyses with the original respondent sample—student enrollment, student race/ethnicity, and the percentage of students eligible for free or reduced-price lunch. Means for each continuous variable for nonresponding and responding schools in the final sample after nonresponse weight

adjustment are shown in tables B-12 through B-14. The difference between the means of responding and nonresponding schools was tested using a *t* test. The *p* values for the *t* tests and the relative bias are presented in the tables.

This set of analyses found a statistically significant difference between responding and nonresponding schools for fourth-grade enrollment (table B-12), with responding schools having fewer fourth-grade students. Also, the relative bias for total enrollment is about 26 percent, which indicates potential bias even though no statistically significant difference was detected. Looking at race/ethnicity (table B-13), there were differences between responding and nonresponding schools in the percentages of Asian or Pacific Islander students and in the Other race/ethnicity category. That is, Asian or Pacific Islander students tended to be underrepresented and students of the Other race/ethnicity category tended to be overrepresented among the responding schools compared with the nonresponding schools. Also, the relative bias for Black and Hispanic students is greater than 10 percent, so these students may be underrepresented by responding schools. No measurable difference was detected for free or reduced-price lunch status (table B-14). The relative bias was about 27 percent for free or reduced-priced lunch status, which indicates potential bias even though no statistically significant difference was detected.

Table B-12. Mean enrollment of nonresponding and responding schools in the U.S. PIRLS final sample before nonresponse weight adjustment: 2006

Student enrollment	Sample schools		Bias	Relative bias	*t* test *p* value
	Responding (standard error)	Nonresponding (standard error)			
Total school	424.35	536.45			
	(20.46)	(44.95)	112.1	0.2642	0.0523
Fourth grade	67.60	95.86			
	(3.72)	(9.44)	28.26	0.4180	0.0211

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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-13. Mean percentage of students in nonresponding and responding schools in the U.S. PIRLS final sample, by race/ethnicity before nonresponse weight adjustment: 2006

Race/ethnicity	Sample schools: Responding (standard error)	Sample schools: Nonresponding (standard error)	Bias	Relative bias	*t* test *p* value
Asian or Pacific Islander	2.48	4.86			
	(0.27)	(0.99)	2.38	0.9597	0.0236
Black, non–Hispanic	15.84	17.52			
	(1.86)	(3.40)	1.68	0.1061	0.6679
Hispanic	15.68	22.29			
	(1.57)	(4.13)	6.61	0.4216	0.1449
American Indian or Alaska Native	2.11	1.92			
	(0.89)	(0.93)	–0.19	–0.0900	0.8830
White, non–Hispanic	53.80	51.29			
	(2.55)	(5.05)	–2.51	–0.0467	0.6613
Other	10.10	2.12			
	(1.78)	(0.62)	–7.98	–0.7901	0.0000

NOTE: Data on race/ethnicity were missing for 12 public schools. 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. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-14. Mean percentage of public school students eligible for free or reduced-price lunch in nonresponding and responding public schools in the U.S. PIRLS final sample before nonresponse weight adjustment: 2006

Students	Sample schools: Responding (standard error)	Sample schools: Nonresponding (standard error)	Bias	Relative bias	*t* test *p* value
Percentage of students eligible for free or reduced-price lunch	33.61	42.72			
	(2.26)	(4.40)	9.11	0.2711	0.0683

NOTE: Free or reduced-price lunch information was available only for public schools. 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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

B.3.4 Continuous Variables—After Nonresponse Weight Adjustment

Means for each continuous variable for nonresponding and responding schools in the final sample after nonresponse weight adjustment are shown in tables B-15 through B-17. The difference between the means of responding and eligible schools was tested using a *t* test. The *p* values for the *t* tests and the relative bias are presented in the tables.

No measurable difference was detected for enrollment (table B-15), race/ethnicity (table B-16), or free or reduced-price lunch status (table B-17). Compared with bias before weight adjustment, the absolute value of the bias after weight adjustment decreased for all school

characteristics. Also, none of the computed relative biases after weight adjustment are greater than 10 percent.

Table B-15. Mean enrollment of full sample and responding schools in the U.S. PIRLS final sample after nonresponse weight adjustment: 2006

Student enrollment	Sample schools: Responding (standard error)	Sample schools: Full sample (standard error)	Bias	Relative bias	*t* test *p* value
Total school	431.42	437.01			
	(20.23)	(19.20)	5.59	0.0123	0.6981
Fourth grade	68.63	70.79			
	(3.73)	(3.57)	2.16	0.0315	0.5615

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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-16. Mean percentage of students in full sample and responding schools in the U.S. PIRLS final sample after nonresponse weight adjustment, by race/ethnicity: 2006

Race/ethnicity	Sample schools: Responding (standard error)	Sample schools: Full sample (standard error)	Bias	Relative bias	*t* test *p* value
Asian or Pacific Islander	2.56	2.81			
	(0.28)	(0.27)	0.25	0.0977	0.1318
Black, non-Hispanic	16.11	16.07			
	(1.93)	(1.66)	–0.04	–0.0025	0.9546
Hispanic	16.14	16.60			
	(1.58)	(1.44)	0.46	0.0285	0.5080
American Indian or Alaska Native	1.90	2.08			
	(0.77)	(0.78)	0.18	0.0947	0.1727
White, non-Hispanic	53.90	53.45			
	(2.57)	(2.29)	–0.45	–0.0083	0.6385
Other	9.39	8.99			
	(1.63)	(1.54)	–0.40	–0.0426	0.2107

NOTE: Data on race/ethnicity were missing for 12 public schools. 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. Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

Table B-17. Mean percentage of public school students eligible for free or reduced-price lunch in full sample and responding public schools in the U.S. PIRLS final sample after nonresponse weight adjustment: 2006

Students	Sample schools: Responding (standard error)	Sample schools: Full sample (standard error)	Bias	Relative bias	*t* test *p* value
Percentage of students eligible for free or reduced-price lunch	33.52 (2.32)	34.87 (2.06)	1.35	0.0403	0.1178

NOTE: Free or reduced-price lunch information was available only for public schools. 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: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

B.3.5 Logistic Regression Model

To examine the joint relationship of various characteristics to school nonresponse, the analysis employed a logistic regression model with response status as the binary dependent variable and frame characteristics as predictor variables. The equation was first modeled in SAS using the stepwise procedure using the design weight[4] and an alpha = 0.05 inclusion criteria. That is, predictor variables were added one at a time, and if a predictor variable was significant at the 5 percent level, then it was kept in the model. Public and private schools were modeled together using the variables available for all schools; private schools were considered low-poverty schools. The significant variables in this initial model 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.

The results in table B-18 indicate differences between responding and nonresponding schools in the percentage of Asian or Pacific Islander students, percentage of White students, percentage of American Indian students, and percentage of students in the Northeast Census region. The percentages of Asian or Pacific Islander, White, and American Indian students and students in the Northeast were higher in nonresponding schools than in responding schools. The result for White and American Indian students is inconsistent with the results in table B-13 due to the race/ethnicity variable being defined differently. The SUDAAN estimates, standard errors, test statistics and *p* values are reported in table B-18.

[4] This analysis was only done using the design weight for respondents and nonrespondents before weight adjustment and not using the final weight for respondents after weight adjustment because the model uses response status as the dependent variable.

Table B-18. Logistic regression reduced model parameters in the U.S. PIRLS final sample: 2006

Parameter	Estimate	Standard error	Test for H_0: parameter = 0	*p* value
Intercept	7.30	1.90	3.84	0.0002
Percent Asian or Pacific Islander students	–0.20	0.07	–2.93	0.0038
Category 1: Schools that had fewer than 62 fourth-graders	1.39	0.76	1.82	0.0705
Category 2: Schools containing between 62 and 85 fourth-graders	0.96	0.69	1.39	0.1656
Category 3: Schools containing between 85 and 129 fourth-graders	1.21	0.63	1.93	0.0555
Census Region: Northeast	–2.06	0.84	–2.46	0.0148
Census Region: Southeast	–0.85	0.74	–1.14	0.2543
Census Region: Central	–0.67	0.88	–0.77	0.4431
Percent of eligible students receiving free or reduced-price lunch	–0.03	0.02	–1.65	0.1000
Percent White students	–0.02	0.01	–2.04	0.0421
Percent American Indian students	–0.04	0.02	–2.28	0.0237
Publicly controlled school	–1.36	1.17	–1.16	0.2481
Central city	–0.59	0.91	–0.65	0.5194
Urban fringe/large town	–0.70	0.76	–0.92	0.3570
High poverty status	0.50	0.87	0.57	0.5683
Percent Black students	–0.02	0.01	–1.34	0.1805
Percent Hispanic students	–0.01	0.01	–0.90	0.3710

NOTE: H_0 = null hypothesis.
SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006

The results in table B-19 indicate the results of the full model. Total enrollment is the only variable included in the full model but not in the reduced model. As in the reduced model, there are differences between responding and nonresponding schools in the percentage of Asian or Pacific Islander students, percentage of White students, percentage of American Indian students, and percentage of students in the Northeast Census region. The percentages of Asian or Pacific Islander, White, and American Indian students and students in the Northeast were higher in nonresponding schools than in responding schools. The result for White and American Indian students is inconsistent with the results in table B-13 due to the race/ethnicity variable being defined differently. The SUDAAN estimates, standard errors, test statistics, and *p* values are reported in table B-19.

Table B-19. Logistic regression full model parameters using the U.S. PIRLS final sample: 2006

Parameter	Estimate	Standard error	Test for H_0: parameter = 0	*p* value
Intercept	7.42	1.99	3.73	0.0002
School control, public	–1.34	1.20	–1.12	0.2654
Region—Northeast	–2.06	0.84	–2.45	0.0150
Region—Southeast	–0.84	0.74	–1.13	0.2585
Region—Central	–0.68	0.89	–0.76	0.4455
Central city	–0.58	0.92	–0.63	0.5320
Urban fringe/large Town	–0.69	0.76	–0.91	0.3663
Category 1: Schools that had fewer than 62 fourth-graders	1.31	0.94	1.40	0.1636
Category 2: Schools containing between 62 and 85 fourth-graders	0.91	0.90	1.01	0.3150
Category 3: Schools containing between 85 and 129 fourth-graders	1.18	0.67	1.77	0.0783
Total enrollment	–0.00	0.00	–0.14	0.8898
Percent of students eligible to receive free or reduced-price lunches	–0.03	0.02	–0.62	0.1058
Percent Hispanic	–0.01	0.01	–0.87	0.3866
Percent White	–0.02	0.01	–2.06	0.0411
Percent Black	–0.02	0.01	–1.34	0.1806
Percent American Indian	–0.04	0.02	–2.29	0.0230
Percent Asian or Pacific Islander	–0.20	0.07	2.92	0.0038
High poverty	0.51	0.89	0.57	0.5670

NOTE: H_0 = null hypothesis.
SOURCE: International Association for the Evaluation of Educational Achievement, Progress in International Reading Literacy Study (PIRLS), 2006.

B.3.6 Conclusion

This report presents nonresponse bias analysis results for both the original U.S. PIRLS 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 evidence of some significant nonresponse bias before weight adjustment with respect to total enrollment and the race/ethnicity composition of the responding schools, especially the Asian, Hispanic, and Other categories. However, after weight adjustment, the significant bias is limited to the Asian and Other students with biases of 0.9 and -3.0 percent, respectively. In addition, the relative bias before weight adjustment indicates the potential for bias in school control, community type, Census region, and poverty level. After weight adjustment, the relative bias is about 14 percent for the Northeast Census region and about 27 percent for the Central Census region.

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 with respect to race/ethnicity, school control (private/public), and enrollment. However, after weight adjustment, no significant bias remains. In addition, the relative bias before weight adjustment indicates the potential for bias in community type, Census region, poverty level, and free or reduced-price lunch status. After weight adjustment, the relative bias is about 12 percent for private schools.

Overall, the data suggest that while significant nonresponse bias and large relative biases exist before weight adjustments, the biases are generally reduced after weight adjustment. 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.

B.4 Technical Notes

B.4.1 Description of Variables

Frame characteristics for public schools were taken from the 2003–04 Common Core of Data and, for private schools, from a preliminary version of the 2003–04 Private School Universe Survey.

B.4.1.1 Race/Ethnicity

Students' race/ethnicity was obtained through responses to a two-part question. Students were first asked whether they were Hispanic or Latino and then whether they were members of the following racial groups: American Indian/Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, or White. Multiple responses were allowed. Other includes respondent answers that do not fall into any of the above categories. The answer to the Hispanic/Latino question took precedence. That is, if the response was affirmative to this question, then the respondent was coded as Hispanic/Latino. Those who did not claim to be Hispanic/Latino were coded into one of the five remaining race/ethnic groups.

B.4.1.2 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.4.1.3 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.4.1.4 Poverty Level in Public Schools (Percentage of Students Eligible for Free or Reduced-price Lunch)

Information on poverty level in public schools 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. For the analyses included in this report, responses were grouped into high poverty—schools in which 50 percent or more of students were eligible—and low poverty—schools in which less than 50 percent of students were eligible. Analyses also used the original, continuous version of this variable, namely, the proportion of students eligible to receive free or reduced-price lunch.

All private schools are 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 is not available on the frame.

B.4.2 Statistical Procedures

B.4.2.1 Sampling

The sampling frame for the U.S. school sample for PIRLS 2006 was constructed using data from the 2002–03 Common Core of Data and preliminary data from the 2003–04 Private School Universe Survey. To be consistent with the sampling design for PIRLS 2001, the frame was divided into two parts as follows: (1) a Metropolitan Statistical Area (MSA) stratum was created that included schools located in the 10 most populous MSAs according to the U.S. Census Bureau; and (2) all schools outside those MSAs were grouped into 451 Primary Sampling Units (PSUs)[5] by sorting on MSA and then by county. PSUs were designed to fit within state boundaries and, where possible, within county and city boundaries. Before the selection process, schools were sorted by state, percentage of racial/ethnic minority students, control of school (public/private), percentage of students eligible for free or reduced-priced

[5] PSUs are the first-level sampling units. They are geographic areas consisting of counties or groups of counties.

lunch, and locale. Within each PSU and MSA stratum, schools were selected on the basis of the number of fourth-grade students in the school, so that larger schools had a higher probability of selection than smaller schools. The final sample included 214 schools; 150 were chosen from PSUs and 64 were selected from the MSA stratum.

B.4.2.2 Substitute Schools

To avoid sample size losses, the PIRLS sampling plan identified, *a priori*, substitute schools for each sampled school. Therefore, if an originally selected school refused to participate in the study, it was possible to replace it with a school that already was identified prior to school sampling. Each originally selected school had two preassigned substitute schools. In general, the school immediately following the originally selected school on the ordered sampling frame and the one immediately preceding it were designated as substitute schools. Substitute schools always belonged to the same explicit stratum, although they could come from different implicit strata if the originally selected school was either the first or last school of an implicit stratum.

B.4.2.3 Weighting

Records from the sample schools and students were assigned sampling weights to adjust for over- or underrepresentation from a particular group. The use of sampling 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 described in B.4.2.1.

B.4.2.4 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 fourth-grade students from the 2005–06 school year was just one of many possible samples that could have been selected. Therefore, estimates produced from the PIRLS 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 because of sampling when estimating a statistic, and is often included in reports containing estimates from survey data. The approach used for calculating sampling variances in PIRLS 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.4.2.5 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.4.2.6 *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.4.2.7 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.4.2.8 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 $X_{1i}, X_{2i} \ldots, X_{pi}$ are *p* auxiliary variables associated with the *i*th sampled beneficiary, and $\beta_0, \beta_1, \ldots, \beta_p$ 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 School Recruitment Letter

[DATE]

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

Dear [PRINCIPAL]:

I am writing to let you know that <one/two> of the fourth-grade classrooms in your school <has/have> been selected to participate in the Progress in International Reading Literacy Study (PIRLS 2006). PIRLS 2006 is an international study designed to provide internationally comparable information about the reading literacy of fourth grade students. Students in over 40 countries, including the United States, are participating in PIRLS 2006. The study 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.

We will conduct PIRLS in the spring of 2006. We know that your calendars are busy and we will do our best to be flexible in scheduling a day and time that works best for you. The test consists of two reading passages (about 40 minutes each) and a brief questionnaire. As a token of our appreciation, students in the selected classroom<s> will each receive a book of their choice as well as a certificate and medal denoting that they represented the U.S. in this study. We will also provide an honorarium to the teacher or staff person designated to help us coordinate the study. Your participation is vitally important if we are to achieve the response rates required by the international standards governing the study and we will work with you in anyway possible to facilitate the participation of selected 4th grade students in your school.

PIRLS 2006 provides a unique opportunity to evaluate the reading knowledge and skills of fourth grade 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 reading literacy, and to examine the factors associated with the acquisition of reading literacy among young children. 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. More information about the study 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. Laurence Ogle at NCES for questions about PIRLS 2006 at (202) 502-7426 or visit the PIRLS 2006 website: http://nces.ed.gov/surveys/pirls.

We look forward to working with you on this important study.

Sincerely,

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

D.2 District Notification Letter

[DATE]

[NAME]
[DISTRICT NAME]
[ADDRESS 1]
[ADDRESS 2]
[CITY, STATE ZIP]

Dear [DISTRICT SUPERINTENDENT]:

I am writing to inform your school district about an upcoming international study: the Progress in International Reading Literacy Study (PIRLS 2006). PIRLS 2006 is designed to provide internationally comparable information about the reading literacy of fourth grade students. Students in over 40 countries, including the United States, are participating in PIRLS 2006. 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. PIRLS 2006 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.

PIRLS 2006 provides a unique opportunity to evaluate the reading knowledge and skills of fourth grade 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 reading literacy, and to examine the factors associated with the acquisition of reading literacy among young children.

We ask you to support the participation of schools in your district in the PIRLS 2006 study. Information about districts, students, and schools sampled for PIRLS 2006 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 PIRLS 2006. In the next few days, we will contact the following schools in your district which have been selected for PIRLS 2006: [LIST ORIGINAL SCHOOLS HERE]

We may also contact: [LIST SUBSTITUTE SCHOOLS HERE]

We will be conducting the PIRLS study in the spring 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 PIRLS 2006 or your district's participation in the study, please call Dr. Patricia J. Green at RTI at (877) 225-0771. For more information about PIRLS 2006, you may contact Dr. Laurence Ogle at NCES at (202) 502-7426 or visit the PIRLS website: http://nces.ed.gov/surveys/pirls/.

Thank you for your support of PIRLS 2006.

Sincerely,

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

D.3 PIRLS 2006 Fact Sheet

PIRLS

Progress in International Reading Literacy Study

This fact sheet answers some important questions about the Progress in International Reading Literacy Study (PIRLS 2006).

What is PIRLS 2006?

PIRLS 2006 is an internationally standardized assessment designed by participating countries and administered on a planned 5-year cycle. The centerpiece of the study is a student reading literacy assessment, which will be administered to fourth-graders to assess their reading achievement. PIRLS 2006 will also use questionnaires completed by the students tested, their teachers, and their school principals to gather information about the factors associated with the development of reading literacy and about the larger contexts in which children live and learn.

Why is participation important?

This study investigates reading literacy and the factors associated with its acquisition in countries around the world. 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 compare with their peers around the world has become an even more prominent issue than ever before. Beyond just simple comparisons, understanding what other nations are doing to further the educational achievement of their populations has also become increasingly more important.

When will the test administration be conducted?

The test administration period is between January 23 and March 24, 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?

Participation in PIRLS 2006 is entirely voluntary, but the participation of every school, student, and teacher is important to ensure the completeness and accuracy of results. Development of national results, and inclusion in the international comparisons, depend on a high response rate.

How many countries participate?

In 2001, the United States was one of 35 countries to participate in the first assessment of PIRLS. Over 40 countries are scheduled to participate in PIRLS 2006, including: Australia, Austria, Belarus, Belgium, Bulgaria, Canada, China, Chinese Taipei (Taiwan), Denmark, England, France, Germany, Hong Kong, Hungary, Iceland, Indonesia, Iran, Israel, Italy, Kuwait, Latvia, Lithuania, Luxembourg, Macedonia, Morocco, Netherlands, New Zealand, Nicaragua, Norway, Poland, Qatar, Romania, Russian Federation, Scotland, Singapore, Slovak Republic, Slovenia, South Africa, Spain, Sweden, United States, and Zimbabwe.

Are there released test items we can review?

Yes, you can see all of the released reading passages, test items, and the scoring guide for the PIRLS 2001 assessment by going to http://isc.bc.edu/pirls2001i/pdf/P1_IR_C.pdf

How many schools, students, and staff will be selected?

The main study of PIRLS 2006 will consist of approximately 180 schools. One, two, or three fourth grade classes will be selected from each school to participate in the study. The primary teacher of each selected class will be asked to complete a teacher survey, and one school administrator will be asked to complete the school questionnaire.

Will the responses of participants be kept confidential?

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 assessment and questionnaires have been designed to keep burden to a minimum. The test and survey will take approximately 2 hours to complete. The school administrator and teacher questionnaires will each take about 30 minutes.

We recognize that instructional time is valuable and want to add to students' learning experience. Each participating student will make a selection from a Scholastic book list as a token of our appreciation. In addition, we offer an honorarium of $75 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 fourth grade reading literacy, the factors associated with reading achievement, and U.S. students compare with fourth graders 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 PIRLS 2006?

PIRLS 2006 is endorsed by the following organizations: American Federation of Teachers, American Association of School Administrators, the 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 PIRLS 2006?

PIRLS is sponsored internationally by the International Association for the Evaluation of Educational Achievement (IEA), and sponsored in the U.S. by the National Center for Education Statistics (NCES), U.S. Department of Education. NCES has contracted RTI International (RTI), 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 Dr. Patricia J. Green or Cathy Forstner at RTI toll-free at (877) 225-0771 between 9AM and 5PM Eastern time, or Dr. Laurence Ogle at the National Center for Education Statistics at (202) 502-7426. You may also visit the PIRLS website: http://nces.ed.gov/surveys/pirls/.

D.4 Implicit Consent Letter and Form

U.S. DEPARTMENT OF EDUCATION
INSTITUTE OF EDUCATION SCIENCES

NATIONAL CENTER FOR EDUCATION STATISTICS

Dear Parent or Guardian:

We are pleased to inform you that your child's class has been selected to participate in an important international study called the Progress in International Reading Literacy Study, or PIRLS 2006. PIRLS 2006 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 literacy in reading, and to compare the reading ability of students in the United States to students in over 40 other countries over time. 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 child will be asked to spend about two hours to complete the questionnaire and test along with approximately 25-50 other students in his/her school. Your child is one of about 4,500 fourth-graders from about 180 schools participating in PIRLS during the spring term of 2006.

Participation is voluntary. You or your child may withdraw from the study at any point. There is no penalty if you or your child decides not to participate. However, we do need your help. Your child was selected to represent many others. His/her responses are necessary to make the results of this important study accurate and timely. Your child may choose not to answer any question. There are no risks or direct benefits to your son or daughter from taking part in the study. Results of this study may help all students in the future. Students will select a free book of their choice from a Scholastic Books list following their participation in the study.

Information about districts, students, and schools sampled for PIRLS 2006 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.

If for any reason you object to your son's or daughter's participation, you may simply deny permission. If you do not want your son or daughter to participate, please fill out the enclosed form and return it to your child's school in the enclosed envelope as soon as possible. If you are willing to allow your son or daughter to participate, you do not need to return this form.

The enclosed brochure gives more information about the study. If you have any questions about PIRLS 2006 or your child's participation in the study, please call Cathy Forstner at RTI, toll-free, at (877) 225-0771 between 9 AM and 5 PM Eastern time, Monday through Friday. If you have questions about your child's rights as a study participant, you may call RTI's Office for Research Protection in Durham, NC, 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 research.

Sincerely,

Val Plisko

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

WASHINGTON, D.C. 20006–

PROGRESS IN INTERNATIONAL READING LITERACY STUDY (PIRLS)
PERMISSION FORM

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

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

I DO NOT GRANT PERMISSION for my child, _____, to participate in the Progress in International Reading Literacy Study (PIRLS).

_____

(Signature of parent or guardian)

Date of signature: _____

(_____)_____
Area code Telephone number

PLEASE PRINT:

Student name: _____

School name: _____

FOR OFFICE USE ONLY:

Student ID: _____

D.5 Explicit Consent Letter and Form

Dear Parent or Guardian:

We are pleased to inform you that your child's class has been selected to participate in an important international study called the Progress in International Reading Literacy Study, or PIRLS 2006. PIRLS 2006 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 literacy in reading, and to compare the reading ability of students in the United States to students in over 40 other countries over time. 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 child will be asked to spend about two hours to complete the questionnaire and test along with approximately 25-50 other students in his/her school. Your child is one of about 4,500 fourth-graders from about 180 schools participating in PIRLS during the spring term of 2006.

Participation is voluntary. You or your child may withdraw from the study at any point. There is no penalty if you or your child decides not to participate. However, we do need your help. Your child was selected to represent many others. His/her responses are necessary to make the results of this important study accurate and timely. Your child may choose to not answer any question. There are no risks or other direct benefits to your son or daughter from taking part in the study. Results of this study may help all students in the future. Students will select a free book of their choice from a Scholastic Books list following their participation in the study.

Information about districts, students, and schools sampled for PIRLS 2006 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.

Please take a moment in the next day or two to fill out the enclosed form and return it to your child's school in the enclosed envelope. We will need to know whether you will allow your son or daughter to participate in this study. We cannot allow your child to participate without your written consent.

The enclosed brochure gives more information about the study. If you have any questions about PIRLS 2006 or your child's participation in the study, please call Cathy Forstner at RTI, toll-free, at (877) 225-0771 between 9 AM and 5 PM Eastern time, Monday through Friday. If you have questions about your child's rights as a study participant, you may call RTI's Office for Research Protection at 1-866-214-2043 in Durham, NC. Both Ms. Forstner and staff from the Office for Research Protection can be reached at: RTI, P.O. Box 12194, Research Triangle Park, NC 27709.

We thank you in advance for your cooperation in this important research.

Sincerely,

Val Plisko
Associate Commissioner
National Center for Education Statistics

Progress in International Reading Literacy Study (PIRLS) Permission Form

Please check the line that indicates your decision about your child's participation in the study. Please check only one option and fill out your child's name, and your signature, phone number and school name at the bottom of the form.

PLEASE RETURN THIS FORM TO YOUR CHILD'S SCHOOL AS SOON AS POSSIBLE. WE HAVE ENCLOSED AN ENVELOPE ADDRESSED TO YOUR CHILD'S TEACHER.

Please check one:

_____ **I GIVE PERMISSION** for my child, _____, to participate in the Progress In International Reading Literacy Study (PIRLS).

_____ **I DO NOT GIVE PERMISSION** for my child, _____, to participate in the Progress In International Reading Literacy Study (PIRLS).

_____

(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 Tables

Table E-1. Practice paper results by evaluator and item

Antarctica

Item number	Evaluator 1 Number correct	Evaluator 1 Number papers	Evaluator 2 Number correct	Evaluator 2 Number papers	Evaluator 3 Number correct	Evaluator 3 Number papers	Evaluator 4 Number correct	Evaluator 4 Number papers	Evaluator 5 Number correct	Evaluator 5 Number papers	Evaluator 6 Number correct	Evaluator 6 Number papers	Evaluator 7 Number correct	Evaluator 7 Number papers	Total Number correct	Total Number papers	Percent by Item
Total	55	60	59	60	57	60	55	60	59	60	60	60	56	60	401	420	95.5
4	11	11	11	11	11	11	9	11	11	11	11	11	11	11	75	77	97.4
7	15	16	16	16	15	16	15	16	15	16	16	16	14	16	106	112	94.6
8	10	11	11	11	10	11	10	11	11	11	11	11	10	11	73	77	94.8
9	11	12	11	12	12	12	11	12	12	12	12	12	11	12	80	84	95.2
11	8	10	10	10	9	10	10	10	10	10	10	10	10	10	67	70	95.7
Percent by evaluator		91.7		98.3		95.0		91.7		98.3		100.0		93.3		95.5	

Leonardo

Item number	Evaluator 1 Number correct	Evaluator 1 Number papers	Evaluator 2 Number correct	Evaluator 2 Number papers	Evaluator 3 Number correct	Evaluator 3 Number papers	Evaluator 4 Number correct	Evaluator 4 Number papers	Evaluator 5 Number correct	Evaluator 5 Number papers	Evaluator 6 Number correct	Evaluator 6 Number papers	Evaluator 7 Number correct	Evaluator 7 Number papers	Total Number correct	Total Number papers	Percent by Item
Total	71	77	74	77	73	77	65	77	73	77	74	77	67	77	497	539	92.2
3	7	7	7	7	7	7	6	7	7	7	7	7	7	7	48	49	98.0
4	20	22	22	22	21	22	12	22	22	22	22	22	15	22	134	154	87.0
6	9	10	9	10	9	10	10	10	9	10	9	10	9	10	64	70	91.4
8	10	10	10	10	10	10	10	10	10	10	10	10	10	10	70	70	100.0
10	13	14	14	14	13	14	13	14	14	14	14	14	13	14	94	98	95.9
12	12	14	12	14	13	14	14	14	11	14	12	14	13	14	87	98	88.8
Percent by evaluator		92.2		96.1		94.8		84.4		94.8		96.1		87.0		92.2	

Table E-1. Practice Paper Results by Evaluator and Item—Continued

Day Hiking

Item number	Evaluator 1		2		3		4		5		6		7		Total		Percent by Item
	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	
Total	47	51	49	51	49	51	48	51	49	51	46	51	46	51	334	357	93.6
2	7	7	7	7	7	7	7	7	7	7	7	7	7	7	49	49	100.0
3	10	10	10	10	10	10	10	10	10	10	10	10	10	10	70	70	100.0
8	12	12	12	12	12	12	11	12	12	12	11	12	11	12	81	84	96.4
11	9	11	10	11	9	11	9	11	10	11	8	11	9	11	64	77	83.1
12	9	11	10	11	11	11	11	11	10	11	10	11	9	11	70	77	90.9
Percent by evaluator		92.2		96.1		96.1		94.1		96.1		90.2		90.2		93.6	

Sharks

Item number	Evaluator 1		2		3		4		5		6		7		Total		Percent by Item
	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	
Total	55	58	54	58	56	58	52	58	55	58	55	58	52	58	379	406	93.3
1	10	10	10	10	10	10	10	10	10	10	10	10	10	10	70	70	100.0
2	11	11	11	11	11	11	11	11	11	11	11	11	10	11	76	77	98.7
5	8	9	9	9	9	9	7	9	8	9	9	9	8	9	58	63	92.1
7	12	12	12	12	12	12	10	12	11	12	12	12	12	12	81	84	96.4
10	7	8	5	8	6	8	7	8	7	8	6	8	4	8	42	56	75.0
12	7	8	7	8	8	8	7	8	8	8	7	8	8	8	52	56	92.9
Percent by evaluator		94.8		93.1		96.6		89.7		94.8		94.8		89.7		93.3	

Table E-1. Practice Paper Results by Evaluator and Item—Continued

Fly, Eagle

Item number	Evaluator 1		2		3		4		5		6		7		Total		Percent by Item
	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	
Total	50	51	45	51	44	51	40	51	46	51	49	51	43	51	317	357	88.8
5	7	8	6	8	8	8	7	8	8	8	8	8	7	8	51	56	91.1
7	14	14	11	14	10	14	9	14	11	14	12	14	12	14	79	98	80.6
9	12	12	11	12	11	12	8	12	11	12	12	12	9	12	74	84	88.1
10	6	6	6	6	6	6	6	6	6	6	6	6	6	6	42	42	100.0
12	11	11	11	11	9	11	10	11	10	11	11	11	9	11	71	77	92.2
Percent by evaluator		98.0		88.2		86.3		78.4		90.2		96.1		84.3		88.8	

Clay

Item number	Evaluator 1		2		3		4		5		6		7		Total		Percent by Item
	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	
Total	64	67	67	67	65	67	60	67	64	67	65	67	59	67	444	469	94.7
2	8	9	9	9	9	9	9	9	9	9	9	9	7	9	60	63	95.2
3	9	9	9	9	9	9	8	9	9	9	9	9	8	9	61	63	96.8
6	9	9	9	9	9	9	9	9	9	9	9	9	8	9	62	63	98.4
8	12	13	13	13	13	13	13	13	13	13	13	13	13	13	90	91	98.9
10	14	15	15	15	15	15	11	15	13	15	14	15	13	15	95	105	90.5
11	12	12	12	12	10	12	10	12	11	12	11	12	10	12	76	84	90.5
Percent by evaluator		95.5		100.0		97.0		89.6		95.5		97.0		88.1		94.7	

Table E-1. Practice Paper Results by Evaluator and Item—Continued

Flowers

Item number	Evaluator 1		2		3		4		5		6		7		Total		Percent by Item
	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	
Total	56	62	60	62	58	62	†	†	60	62	57	62	54	62	345	372	92.7
6	9	9	9	9	9	9	†	†	9	9	9	9	9	9	54	54	100.0
7	10	11	11	11	11	11	†	†	11	11	10	11	10	11	63	66	95.5
8	8	8	7	8	8	8	†	†	8	8	8	8	7	8	46	48	95.8
9	9	9	8	9	9	9	†	†	8	9	8	9	8	9	50	54	92.6
10	8	8	8	8	8	8	†	†	8	8	8	8	8	8	48	48	100.0
12	12	17	17	17	13	17	†	†	16	17	14	17	12	17	84	102	82.4
Percent by evaluator		90.3		96.8		93.5		†		96.8		91.9		87.1		92.7	

Shiny Straw

Item number	Evaluator 1		2		3		4		5		6		7		Total		Percent by Item
	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	
Total	69	74	71	74	70	74	†	†	70	74	70	74	63	74	413	444	93.0
3	10	10	10	10	10	10	†	†	10	10	10	10	9	10	59	60	98.3
9	11	12	11	12	11	12	†	†	12	12	12	12	10	12	67	72	93.1
10	11	11	11	11	10	11	†	†	10	11	11	11	9	11	62	66	93.9
12	13	13	13	13	13	13	†	†	13	13	11	13	12	13	75	78	96.2
13	11	13	12	13	13	13	†	†	11	13	13	13	9	13	69	78	88.5
14	13	15	14	15	13	15	†	†	14	15	13	15	14	15	81	90	90.0
Percent by evaluator		93.2		95.9		94.6		†		94.6		94.6		85.1		93.0	

Table E-1. Practice Paper Results by Evaluator and Item—Continued

Unbelievable

Item number	Evaluator 1		2		3		4		5		6		7		Total		Percent by Item
	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	
Total	40	46	40	46	40	46	†	†	41	46	43	46	40	46	244	276	88.4
6	6	7	7	7	6	7	†	†	7	7	6	7	7	7	39	42	92.9
8	9	11	10	11	7	11	†	†	9	11	10	11	9	11	54	66	81.8
10	6	7	6	7	7	7	†	†	7	7	7	7	7	7	40	42	95.2
11	12	13	11	13	12	13	†	†	11	13	12	13	9	13	67	78	85.9
12	7	8	6	8	8	8	†	†	7	8	8	8	8	8	44	48	91.7
Percent by evaluator		87.0		87.0		87.0		†		89.1		93.5		87.0		88.4	

Food Search

Item number	Evaluator 1		2		3		4		5		6		7		Total		Percent by Item
	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	
Total	73	79	70	79	76	79	†	†	75	79	76	79	72	79	442	474	93.2
5	10	11	9	11	11	11	†	†	11	11	11	11	10	11	62	66	93.9
7	15	15	13	15	15	15	†	†	15	15	15	15	14	15	87	90	96.7
9	8	8	8	8	8	8	†	†	8	8	8	8	8	8	48	48	100.0
10	10	10	10	10	10	10	†	†	10	10	10	10	10	10	60	60	100.0
12	9	10	9	10	9	10	†	†	10	10	10	10	9	10	56	60	93.3
13	12	12	12	12	12	12	†	†	11	12	12	12	12	12	71	72	98.6
15	9	13	9	13	11	13	†	†	10	13	10	13	9	13	58	78	74.4
Percent by evaluator		92.4		88.6		96.2		†		94.9		96.2		91.1		93.2	

Table E-1. Practice Paper Results by Evaluator and Item—Continued

	Evaluator														Total	
	1		2		3		4		5		6		7			
	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers	Number correct	Number papers
Total	580	625	589	625	588	625	320	364	592	625	595	625	552	625	3,816	4,114
Percent by evaluator		92.8		94.2		94.1		87.9		94.7		95.2		88.3		92.8

† Not applicable. Evaluator 4 did not score Leonardo, Clay, Flowers, Shiny Straw, Unbelievable, or Food Search.

Table E-2. Back reading results by evaluator and item

Antarctica

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	69	70	91	91	98	98	106	112	103	105	123	126	98	105	688	707	97.3
1	10	10	13	13	14	14	16	16	14	15	18	18	15	15	100	101	99.0
3	10	10	13	13	14	14	16	16	15	15	18	18	15	15	101	101	100.0
4	10	10	13	13	14	14	16	16	15	15	18	18	15	15	101	101	100.0
7	10	10	13	13	14	14	16	16	15	15	18	18	15	15	101	101	100.0
8	10	10	13	13	14	14	14	16	14	15	17	18	13	15	95	101	94.1
9	10	10	13	13	14	14	13	16	15	15	18	18	13	15	96	101	95.0
11	9	10	13	13	14	14	15	16	15	15	16	18	12	15	94	101	93.1
Agreement by evaluator	98.6		100.0		100.0		94.6		98.1		97.6		93.3		97.3		

Leonardo

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	58	60	89	90	83	90	†	†	105	108	104	108	104	108	543	564	96.3
3	10	10	15	15	15	15	†	†	18	18	18	18	18	18	94	94	100.0
4	10	10	14	15	14	15	†	†	17	18	17	18	17	18	89	94	94.7
6	9	10	15	15	12	15	†	†	17	18	17	18	17	18	87	94	92.6
8	10	10	15	15	14	15	†	†	18	18	18	18	17	18	92	94	97.9
10	10	10	15	15	13	15	†	†	17	18	17	18	17	18	89	94	94.7
12	9	10	15	15	15	15	†	†	18	18	17	18	18	18	92	94	97.9
Agreement by evaluator	96.7		98.9		92.2		†		97.2		96.3		96.3		96.3		

Table E-2. Back reading results by evaluator and item—Continued

Clay

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	69	70	255	259	253	259	†	†	250	252	258	259	241	245	1,326	1,344	98.7
1	10	10	37	37	37	37	†	†	36	36	37	37	35	35	192	192	100.0
2	10	10	34	37	37	37	†	†	35	36	37	37	35	35	188	192	97.9
3	10	10	37	37	36	37	†	†	36	36	36	37	35	35	190	192	99.0
6	10	10	36	37	37	37	†	†	36	36	37	37	35	35	191	192	99.5
8	10	10	37	37	36	37	†	†	36	36	37	37	34	35	190	192	99.0
10	9	10	37	37	33	37	†	†	35	36	37	37	33	35	184	192	95.8
11	10	10	37	37	37	37	†	†	36	36	37	37	34	35	191	192	99.5
Agreement by evaluator	98.6		98.5		97.7		†		99.2		99.6		98.4		98.7		

Sharks

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	60	60	106	108	100	102	81	84	106	108	106	108	113	114	672	684	98.2
1	10	10	17	18	17	17	14	14	17	18	17	18	19	19	111	114	97.4
2	10	10	18	18	17	17	13	14	18	18	18	18	19	19	113	114	99.1
5	10	10	18	18	16	17	14	14	18	18	18	18	19	19	113	114	99.1
7	10	10	18	18	16	17	13	14	18	18	18	18	18	19	111	114	97.4
10	10	10	17	18	17	17	14	14	18	18	18	18	19	19	113	114	99.1
12	10	10	18	18	17	17	13	14	17	18	17	18	19	19	111	114	97.4
Agreement by evaluator	100.0		98.1		98.0		96.4		98.1		98.1		99.1		98.2		

Table E-2. Back reading results by evaluator and item—Continued

Searching

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	192	192	190	192	190	192	†	†	192	192	192	192	188	192	1,144	1,152	99.3
5	24	24	23	24	24	24	†	†	24	24	24	24	23	24	142	144	98.6
7	24	24	24	24	24	24	†	†	24	24	24	24	24	24	144	144	100.0
9	24	24	24	24	23	24	†	†	24	24	24	24	22	24	141	144	97.9
10	24	24	24	24	24	24	†	†	24	24	24	24	24	24	144	144	100.0
11	24	24	24	24	24	24	†	†	24	24	24	24	24	24	144	144	100.0
12	24	24	24	24	24	24	†	†	24	24	24	24	24	24	144	144	100.0
13	24	24	23	24	24	24	†	†	24	24	24	24	24	24	143	144	99.3
15	24	24	24	24	23	24	†	†	24	24	24	24	23	24	142	144	98.6
Agreement by evaluator	100.0		99.0		99.0		†		100.0		100.0		97.9		99.3		

Night

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	119	120	121	126	60	60	†	†	144	144	164	168	123	132	731	750	97.5
5	20	20	21	21	10	10	†	†	24	24	28	28	22	22	125	125	100.0
6	19	20	20	21	10	10	†	†	24	24	28	28	21	22	122	125	97.6
8	20	20	20	21	10	10	†	†	24	24	28	28	20	22	122	125	97.6
10	20	20	19	21	10	10	†	†	24	24	27	28	20	22	120	125	96.0
11	20	20	20	21	10	10	†	†	24	24	27	28	19	22	120	125	96.0
12	20	20	21	21	10	10	†	†	24	24	26	28	21	22	122	125	97.6
Agreement by evaluator	99.2		96.0		100.0		†		100.0		97.6		93.2		97.5		

Table E-2. Back reading results by evaluator and item—Continued

Straw

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	58	60	165	168	166	168	†	†	149	150	146	150	112	114	796	810	98.3
3	10	10	27	28	27	28	†	†	25	25	25	25	19	19	133	135	98.5
9	10	10	28	28	28	28	†	†	24	25	25	25	19	19	134	135	99.3
10	9	10	28	28	28	28	†	†	25	25	25	25	19	19	134	135	99.3
12	9	10	28	28	28	28	†	†	25	25	25	25	18	19	133	135	98.5
13	10	10	26	28	27	28	†	†	25	25	23	25	19	19	130	135	96.3
14	10	10	28	28	28	28	†	†	25	25	23	25	18	19	132	135	97.8
Agreement by evaluator	96.7		98.2		98.8		†		99.3		97.3		98.2		98.3		

Hiking

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	49	50	97	100	156	160	116	120	96	100	101	105	147	150	762	785	97.1
2	10	10	20	20	32	32	23	24	20	20	20	21	30	30	155	157	98.7
3	10	10	20	20	31	32	23	24	20	20	21	21	30	30	155	157	98.7
8	10	10	20	20	31	32	24	24	18	20	20	21	29	30	152	157	96.8
11	10	10	18	20	30	32	22	24	19	20	21	21	28	30	148	157	94.3
12	9	10	19	20	32	32	24	24	19	20	19	21	30	30	152	157	96.8
Agreement by evaluator	98.0		97.0		97.5		96.7		96.0		96.2		98.0		97.1		

Table E-2. Back reading results by evaluator and item—Continued

Eagle

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	47	50	132	135	116	120	†	†	125	130	117	125	134	140	671	700	95.9
5	10	10	27	27	24	24	†	†	26	26	24	25	26	28	137	140	97.9
7	10	10	27	27	24	24	†	†	23	26	24	25	27	28	135	140	96.4
9	9	10	26	27	21	24	†	†	25	26	21	25	25	28	127	140	90.7
10	10	10	27	27	24	24	†	†	26	26	25	25	28	28	140	140	100.0
12	8	10	25	27	23	24	†	†	25	26	23	25	28	28	132	140	94.3
Agreement by evaluator	94.0		97.8		96.7		†		96.2		93.6		95.7		95.9		

Flowers

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	58	60	92	96	141	144	†	†	162	162	147	150	149	150	749	762	98.3
6	9	10	16	16	23	24	†	†	27	27	25	25	25	25	125	127	98.4
7	10	10	14	16	24	24	†	†	27	27	23	25	24	25	122	127	96.1
8	10	10	16	16	24	24	†	†	27	27	24	25	25	25	126	127	99.2
9	10	10	16	16	24	24	†	†	27	27	25	25	25	25	127	127	100.0
10	10	10	16	16	24	24	†	†	27	27	25	25	25	25	127	127	100.0
12	9	10	14	16	22	24	†	†	27	27	25	25	25	25	122	127	96.1
Agreement by evaluator	96.7		95.8		97.9		†		100.0		98.0		99.3		98.3		

Table E-2. Back reading results by evaluator and item—Continued

Item number	Evaluator 1		2		3		4		5		6		7		Total		Agreement by item
	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	Number agree	Number read	
Total	779	792	1,338	1,365	1,363	1,393	303	316	1,432	1,451	1,458	1,491	1,409	1,450	8,082	8,258	
Agreement by evaluator	98.4		98.0		97.8		95.9		98.7		97.8		97.2		97.9		

† Not applicable. Evaluator 4 did not score Leonardo, Clay, Flowers, Shiny Straw, Unbelievable, or Food Search.