

# Progress in International Reading Literacy Study (PIRLS)

Website: <http://nces.ed.gov/surveys/pirls/>

## 1. OVERVIEW

The Progress in International Reading Literacy Study (PIRLS) is a large international comparative study of the reading literacy of fourth-grade students. The study is conducted by the International Association for the Evaluation of Educational Achievement (IEA), with national sponsors in each participating jurisdiction. The National Center for Education Statistics (NCES), in the Institute of Education Sciences at the U.S. Department of Education, is responsible for the implementation of PIRLS in the United States. Reading literacy is one of the most important abilities that students acquire as they progress through their early school years. It is the foundation for learning across all subjects, it can be used for recreation and for personal growth, and it equips young children with the ability to participate fully in their communities and the larger society. Participants in PIRLS include both countries and subnational entities, both of which are referred to as “jurisdictions.” PIRLS focuses on the achievement and reading experiences of children in grades equivalent to fourth grade in the United States. The study includes a written test of reading comprehension and a series of questionnaires focusing on the factors associated with the development of reading literacy. PIRLS was first administered in 2001 to students in 35 jurisdictions and was administered again in 2006 to students in 45 jurisdictions. The next PIRLS is scheduled for 2011.

### Purpose

PIRLS is a carefully constructed reading assessment, consisting of a test of the reading literacy of fourth-grade students and questionnaires to collect information about fourth-grade students’ reading literacy performance. PIRLS has four goals: (1) develop internationally valid instruments for measuring reading literacy suitable for establishing internationally comparable literacy levels in each of the participating jurisdictions; (2) describe on one international scale the literacy profiles of fourth-graders in school in each of the participating jurisdictions; (3) describe the reading habits of fourth-graders in each participating jurisdiction; and (4) identify the home, school, and societal factors associated with the literacy levels and reading habits of fourth-graders in school.

### Components

PIRLS focuses on three aspects of reading literacy: purposes for reading; processes of comprehension; student reading behaviors and engagement. The first two form the basis of the written test of reading comprehension. The student background questionnaire addresses the third aspect.

In PIRLS, purpose for reading refers to the two types of reading that account for most of the reading young students do, both in and out of school: (1) reading for literary experience, and (2) reading to acquire and use information. In the assessment, narrative fiction is used to assess students’ ability to read for literary experience, while a variety of informational texts are used to assess students’ ability to acquire and use information while reading. The PIRLS assessment contains an equal proportion of texts assessing each purpose. Processes of comprehension refer

### PROGRESS IN INTERNATIONAL READING AND LITERACY STUDY:

Three aspects of reading literacy:

- Purpose for reading
- Processes of comprehension
- Reading behaviors and attitudes

Four sets of questionnaires:

- Student questionnaire
- Learning to read (home) survey
- Teacher questionnaire
- School principal questionnaire

to ways in which readers construct meaning from the text. There are four comprehension processes: focusing on and retrieving specific ideas; making inferences; interpreting and integrating ideas and information; and examining or evaluating text features.

**Assessment.** The PIRLS assessment instruments include stories and informational texts at the fourth-grade level collected internationally. Students are asked to engage in a full repertoire of reading skills and strategies, including retrieving and focusing on specific ideas, making simple and more complex inferences, and examining and evaluating text features. The passages are followed by constructed-response and multiple-choice format questions about the text.

In PIRLS 2001, reading passages were printed in some students' assessment booklets, while other students were given the *PIRLS Reader*, a short anthology of a variety of reading texts, in addition to an assessment booklet. Using different booklets allows PIRLS to report results from more assessment items than can fit in one booklet, without making the assessment longer. To provide good coverage of each skill domain, the test items developed required over 5 hours of testing time. However, testing time was kept to 80 minutes for each student by clustering items in 8 blocks distributed across the 10 booklets, (9 student test booklets and the *PIRLS Reader*). Each student completed only one of the booklets. As a consequence, no student received all items, but each item was answered by a representative sample of students.

PIRLS 2006's design was built on PIRLS 2001. To evaluate changes in achievement over time, in 2006 new measuring scales were created in addition to the scale for reading achievement overall. To accommodate these changes, the booklet design expanded to include additional test booklets, and the total assessment time increased. PIRLS 2006 included 10 blocks, consisting of a reading passage and its accompanying questions. Four of the PIRLS 2001 test blocks were kept secure and carried forward for measuring trends in 2006, the six remaining blocks were redesigned. The new materials were added to reflect the broad approaches established for 2001, while refreshing and expanding the range of texts and devising items that brought out the qualities of each passage. The item blocks were then distributed across 13 booklets (including *PIRLS Reader*, a full color, magazine-style booklet) and each student was administered one of the booklets.

**Questionnaires.** Background questionnaires in PIRLS are administered to collect information about students' home and school experiences in learning to read. By

gathering information about children's experiences (together with reading achievement on the PIRLS test), it is possible to identify the factors or combinations of factors that relate to high reading literacy. PIRLS 2001 and PIRLS 2006 administered questionnaires to students, teachers, and school principals. In jurisdictions other than the United States, a parent questionnaire is also administered. Additionally, PIRLS 2006 included a newly constructed curriculum questionnaire that provided information about the national context.

**Student questionnaire.** Each student taking the PIRLS reading assessment completes the student questionnaire. The questionnaire asks about aspects of students' home and school experiences, including instructional experiences and reading for homework, self-perceptions about and attitudes toward reading, out-of-school reading habits, computer use, home literacy resources, and basic demographic information, such as parents' educational level, language spoken at home, and student reading activities.

**Learning to read (home) survey.** The learning to read survey is completed by the parents or primary caregivers of each student taking the PIRLS reading assessment. It addresses child/parent literacy interactions, home literacy resources, parents' reading habits and attitudes, home/school connections, and basic demographic and socioeconomic indicators. This assessment was not administered in the United States in 2001 and 2006.

**Teacher questionnaire.** The reading teacher of each fourth-grade class sampled for PIRLS completes a questionnaire designed to gather information about classroom contexts for developing reading literacy. This questionnaire asks teachers about characteristics of the class tested (such as size, reading levels of the students, and language abilities of the students). It also asks about instructional time, materials and activities for teaching reading and promoting the development of students' reading literacy, and the grouping of students for reading instruction. Questions about classroom resources, assessment practices, and home/school connections are also included. The questionnaire also asks teachers for their views on opportunities for professional development and collaboration with other teachers and for information about their education and training.

**School questionnaire.** The principal of each school sampled for PIRLS responds to the school questionnaire. The questionnaire asks principals about enrollment and other school characteristics (such as where the school is located, resources available in the

surrounding area, and indicators of the socioeconomic background of the student body), characteristics of reading education in the school, instructional time, school resources (such as the availability of instructional materials and staff), home/school connections, and the school climate.

*Curriculum questionnaire.* First used in PIRLS 2006, this questionnaire focused on the nature of the development and implementation of a nationally (or regionally) defined reading curriculum in primary schools within each participating country.

In all, PIRLS takes 1½ to 2 hours of each student's time, including the assessment and background questionnaire.

In addition, system level information was provided by each participating country and published in the PIRLS 2001 Encyclopedia (Mullis et al. 2002) and the PIRLS 2006 Encyclopedia (Kennedy et al. 2007). The encyclopedias provide a description for each participating country of the policies and practices that guide school organization and classroom reading instruction in the lower grades.

### Periodicity

PIRLS is administered once every 5 years, near the end of the school year in each jurisdiction. PIRLS was conducted in 2001 and 2006, and will be administered in the United States and other participating jurisdictions again in 2011.

## 2. USES OF DATA

PIRLS will help educators and policymakers by answering questions such as the following:

- How well do fourth-grade students read?
- How do students in one jurisdiction compare with students in another jurisdiction?
- Do fourth-grade students value and enjoy reading?
- Internationally, how do the reading habits and attitudes of students vary?

## 3. KEY CONCEPTS

*International desired population.* This is the grade or age level that each jurisdiction should address in its sampling activities. The international desired population for PIRLS 2001 was defined as all students

enrolled in the upper of the two adjacent grades that contain the largest proportion of 9-year-olds at the time of testing. For PIRLS 2006, the international desired population was defined as all students enrolled in the grade that represents 4 years of schooling, counting from the 1st year of the International Standard Classification of Education (ISCED) Level 1, providing that the mean age at the time of testing was at least 9.5 years. For most jurisdictions, the target grade was the fourth grade or its national equivalent.

*National desired population.* PIRLS expects all participating jurisdictions to define their national desired population to correspond as closely as possible to the definition of the international desired population. For example, for PIRLS 2001, if the fourth grade was the upper of the two adjacent grades containing the greatest proportion of 9-year-olds in a particular jurisdiction, then students enrolled in fourth grade were the national desired population for that jurisdiction. For PIRLS 2006, if the fourth grade of primary school was the grade that represents 4 years of schooling in a particular jurisdiction (counting from the 1st year of ISCED Level 1), then students enrolled in fourth grade were the national desired population for that jurisdiction.

Although jurisdictions are expected to include all students in the target grade in their definition of the population, sometimes they have to reduce their coverage. Using its national desired population as a basis, each participating jurisdiction has to define its population in operational terms for sampling purposes. Ideally, the national defined population should coincide with the national desired population, although in reality there may be some school types or regions that cannot be included; consequently, the national defined population is usually a very large subset of the national desired population.

*National Research Coordinators (NRCs) and data collection contractor.* Each participating jurisdiction appoints a national research coordinator to monitor national data collection and processing in accordance with international standards. NCES contracts with a data collection firm to draw the samples, work with school coordinators, assemble and print the test booklets, and pack and ship the necessary materials to the sampled schools. The contractor is also responsible for working with school coordinators, translating the test instruments, assembling and printing the test booklets, and packing and shipping the necessary materials to the sampled schools. They are also responsible for arranging the return of the testing materials from the school to the national center, preparing for and implementing the

constructed-response scoring, entering the results into data files, conducting on-site quality assurance observations for a 10 percent sample of schools, and preparing a report on survey activities.

**Reading literacy.** The ability to use printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential. This definition goes beyond simply decoding and comprehending text to include a broad range of information-processing skills that adults use in accomplishing the range of tasks associated with work, home, and community contexts. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers, and for enjoyment. In PIRLS, there is a distinction between reading for literary experience and reading to acquire and use information.

## 4. SURVEY DESIGN

### Target Population

In IEA studies, the target population for all jurisdictions is known as the *international desired population*. The detailed definitions of international desired population for PIRLS 2001 and 2006 are provided in the section of Key Concepts. For both PIRLS 2001 and 2006, the international desired population corresponds to the fourth grade in most jurisdictions, including the United States. This population was chosen because it represents an important transition point in children's development as readers. In most jurisdictions, by the end of fourth grade, children are expected to have learned how to read, and are now reading to learn.

### Sample Design

Using its national desired population as a basis, each participating jurisdiction has to define its population in operational terms for sampling purposes. PIRLS participants are expected to ensure that the national defined population includes at least 95 percent of the national desired population. Exclusions (which should be kept to a minimum) can occur at the school level, within the sampled schools, or at both levels. Because the national desired population is restricted to schools that contain the required grade, schools not containing the target grade are considered to be outside the scope of the sample—not part of the target population.

In each jurisdiction, representative samples of students are selected using a two-stage sampling design. In the first stage, at least 170 schools are selected using probability proportional to size (PPS) sampling. Jurisdictions can incorporate in their sampling design

important reporting variables (for example, urbanicity or school type) as stratification variables. In the second stage, one or two fourth-grade classes are randomly sampled in each school. This results in a sample size of at least 3,750 students in each jurisdiction. Some jurisdictions opt to include more schools and classes, enabling additional analyses, which results in larger sample sizes. In 2006, PIRLS required that all student sample sizes should not be less than 4,000 students.

In the United States in 2001, a nationally representative sample of 3,760 fourth-grade students from 170 schools was selected. The schools were randomly selected first, and then one or two classrooms were randomly selected within each school. In the United States in 2006, a nationally representative sample of 5,190 fourth-grade students from 180 schools was selected. The schools were randomly selected first, and then one or two classrooms were randomly selected within each school.

**First sampling stage.** The sample selection method used for the first sampling stage in PIRLS makes use of a systematic PPS technique. In order to use this method, it is necessary to have some measure of size (MOS) of the sampling units. Ideally, this is the number of sampling elements within the units (e.g., the number of students in the school in the target grade). If this is unavailable, some other highly correlated measure, such as total school enrollment, is used. The schools in each explicit stratum are listed in order of the implicit stratification variables, together with the MOS for each school. Schools are further sorted by MOS within implicit stratification variables. The cumulative MOS is a measure of the size of the population of sampling elements; dividing it by the number of schools to be sampled gives the sampling interval.

The first school is sampled by choosing a random number in the range between 1 and the sampling interval. The school whose cumulative MOS contains the random number is the sampled school. By adding the sampling interval to that first random number, a second school is identified. This process of consistently adding the sampling interval to the previous selection number results in a PPS sample of the required size.

Very large jurisdictions have an opportunity to introduce a preliminary sampling stage before sampling schools. The Russian Federation and the United States avail themselves of this option. In these jurisdictions, the first step is to draw a sample of geographic regions using PPS sampling. Then a sample of schools is drawn from each sampled region. This design is used mostly as a cost reduction measure,

where the construction of a comprehensive list of schools would have been either impossible or prohibitively expensive. Also, the additional sampling stage reduces the dispersion of the school sample, thereby potentially reducing travel costs. Sampling guidelines are put in place to ensure that an adequate number of units will be sampled from this preliminary stage.

**Second sampling stage.** The second sampling stage consists of selecting classrooms within sampled schools. As a rule, one classroom per school is sampled, although some participants opt to sample two classrooms. All classrooms are selected with equal probabilities for all jurisdictions. It is suggested that any classroom smaller than half the specified minimum cluster size be combined with another classroom from the same grade and school.

**Trends in IEA's Reading Literacy Study.** PIRLS jurisdictions that earlier participated in the 1991 IEA Reading Literacy Study had the option of undertaking the Trends in IEA's Reading Literacy Study, which measured trends in reading achievement using IEA's 1991 reading test and student questionnaire. Since the target population for the Trends in IEA's Reading Literacy Study is similar (but not identical) to the PIRLS target population, it is possible to use the PIRLS school sample as the basis for the trend study sample. Accordingly, the sampling plan for the Trends in IEA's Reading Literacy Study is simple: select every second school sampled for PIRLS, and from each of these, sample one additional classroom from the target grade. Since the sample of schools for the Trends in IEA's Reading Literacy Study is essentially a subsample of the PIRLS sample of schools, most of the required sampling tasks are carried out during the PIRLS school sampling.

### Assessment Design

The PIRLS International Study Center is responsible for the design, development, and implementation of the study—including developing the instruments and survey procedures, ensuring quality in data collection, and analyzing and reporting the study results. The PIRLS Reading Development Group contributes to the framework and reading test. Committee members review various drafts of the framework and assessment blocks, and review and endorse the final reading test. The PIRLS Questionnaire Development Group, comprising representatives from nine countries, helps develop the PIRLS questionnaires (including writing items and reviewing drafts of all questionnaires).

**Development of framework and questions.** At the heart of the PIRLS assessment is the definition of

reading literacy established by the Reading Development Group and refined by National Research Coordinators. The PIRLS definition of reading literacy builds on the definition used in the 1991 IEA study, but elaborates on that definition by making specific reference to reading by children.

In accordance with the framework, the passages in the reading test are authentic texts drawn from children's storybooks and informational sources. Submitted and reviewed by the PIRLS jurisdictions, the passages represent a range of types of literary and informational texts. The literary passages include realistic stories and traditional tales, while the informational texts include chronological and nonchronological articles, biographical articles, and informational leaflets.

Two item formats are used to assess children's reading literacy—multiple-choice and constructed-response. Each type of item is used to assess both reading purposes and all four reading processes.

**Matrix sampling.** PIRLS has ambitious goals for covering the domain of reading literacy. The Reading Development Group felt that at least eight passages and items (four for each reading purpose) were needed to provide a valid and reliable measure of reading achievement. Since it would not be possible to administer the entire test to any one student, PIRLS used a matrix sampling technique to distribute the assessment material among students, yet retain linkages necessary for scaling the achievement data.

In PIRLS 2001, assessment material was divided into 40-minute "blocks," each comprised of a passage (a story or article) and items representing at least 15 score points. There were eight such blocks, four for each reading purpose. The eight assessment blocks were distributed across 10 test booklets, and each student completed one booklet in an 80-minute testing session. Each booklet contained two blocks—two literary, two informational, or one of each—and most blocks appeared in three booklets. One of the 10 booklets was the *PIRLS Reader*, a color booklet containing two reading passages; the test items for it were located in a separate booklet. The two blocks for the *Reader* appeared only in that booklet. The distribution of blocks across booklets "links" the booklets to enable the achievement data to be scaled using Item Response Theory (IRT) methods.

The new material developed for PIRLS 2006 was combined with the four secure blocks retained from the 2001 assessment, providing an overall assessment that would allow the calculation of trends over 5 years. The PIRLS 2006 reading assessment was comprised of 13 booklets, one of which was administered to each

student. Each booklet contained two blocks, comprised of a story or article followed by a series of questions pertaining to the text passage. In 2006, there were 10 blocks in total (5 for each reading purpose), which were systematically rotated throughout the booklets. As in 2001, the two blocks for the *Reader* appeared only in that booklet.

### Data Collection and Processing

**Reference dates.** PIRLS is administered near the end of the school year in each jurisdiction. For PIRLS 2001, in jurisdictions in the Northern Hemisphere (where the school year typically ends in May or June), the assessment was conducted in April, May, or June 2001.

In the PIRLS 2006, jurisdictions in the Northern Hemisphere conducted the assessment between March and May 2006. In the United States, data collection began slightly earlier and ended in early June. In the Southern Hemisphere, the school year typically ends in November or December; in these jurisdictions, the assessment was conducted in October or November in 2001 and in October and November in 2005.

**Data collection.** Each jurisdiction is responsible for carrying out all aspects of the data collection, using standardized procedures developed for the study. Manuals provide explicit instructions to the NRCs and their staff members on all aspects of the data collection—from contacting sampled schools to packing and shipping materials to the IEA Data Processing Center for processing and verification. Manuals are also prepared for test administrators and for individuals in the sampled schools who work with the national centers to arrange for the data collection within the schools. These manuals address all aspects of the assessment administration within schools (including test security, distribution of booklets, timing and conduct of the testing session, and returning materials to the national center).

The PIRLS International Study Center places great emphasis on monitoring the quality of the PIRLS data collection. In particular, the Study Center implements an international program of site visits, whereby international Quality Control Monitors (QCMs) visit a sample of 15 schools in each jurisdiction and observe the test administration. In addition to the international program, NRCs are also expected to organize an independent national quality control program based upon the international model. The latter program requires national QCMs to document data collection activities in their jurisdiction. The national QCMs visit a random sample of 10 percent of the schools (in addition to those visited by the international QCMs)

and monitor the testing sessions—recording their observations for later analysis.

**Editing.** To ensure the availability of comparable, high-quality data for analysis, PIRLS takes rigorous quality control steps to create the international database. PIRLS prepares manuals and software for jurisdictions to use in creating and checking their data files, so that the information will be in a standardized international format before being forwarded to the IEA Data Processing Center (DPC) in Hamburg for creation of the international database. Upon arrival at the DPC, the data undergo an exhaustive cleaning process involving several iterative steps and procedures designed to identify, document, and correct deviations from the international instruments, file structures, and coding schemes. The process also emphasizes consistency of information within national datasets and appropriate linking among the student, parent, teacher, and school data files.

Throughout the process, the data are checked and double-checked by the IEA Data Processing Center, the International Study Center, and the national centers. The national centers are contacted regularly and given multiple opportunities to review the data for their jurisdictions. In conjunction with the IEA Data Processing Center, the International Study Center reviews item statistics for each cognitive item in each jurisdiction to identify poorly performing items. In general, the items exhibit very good psychometric properties in all jurisdictions.

### Estimation Methods

**Weighting.** Sampling weights are calculated according to a three-step procedure involving selection probabilities for schools, classrooms, and students.

**School weight.** The first step consists of calculating a school weight, which also incorporates weighting factors from any additional front-end sampling stages, such as districts or regions. A school-level participation adjustment is then made to the school weight to compensate for any sampled schools that do not participate. This adjustment is calculated independently for each explicit stratum.

The PIRLS sample design requires that school selection probabilities be proportional to the school size, defined as enrollment in the target grade. For jurisdictions with a preliminary sampling stage (such as the United States and the Russian Federation), the basic first-stage weight also incorporates the probability of selection in this preliminary stage. The first-stage weight in such cases is simply the product of the “region” weight and the first-stage weight.

In some jurisdictions, schools are selected with equal probabilities. This generally occurs when a large sampling ratio is used. Also, in some jurisdictions, explicit or implicit strata are defined to deal with very large schools or small schools. Equal probability sampling is necessary in these strata.

First-stage weights are calculated for all sampled and replacement schools that participate. A school-level participation adjustment is required to compensate for those schools that are sampled but do not participate and, hence, are not replaced. Sampled schools that are found to be ineligible are removed from the calculation of this adjustment. The school-level participation adjustment is calculated separately for each explicit stratum.

*Classroom weight.* In the second step, a classroom weight reflecting the probability of the sampled classroom(s) being selected from all the classrooms in the school at the target grade level is calculated. All classrooms are sampled with equal probability. No classroom-level participation adjustment is necessary, since in most cases a single classroom is sampled in each school. If a school agrees to take part in the study, but the classroom refuses to participate, adjustment for nonparticipation is made at the school level. If one of two selected classrooms in a school does not participate, then the classroom weight is calculated as though a single classroom has been selected in the first place. The classroom weight is calculated independently for each school.

*Student weight.* Because intact classrooms are sampled in PIRLS, each student in the sampled classrooms is certain of selection, so the base student weight is 1.0. However, as a third and final step, a nonparticipation adjustment is made to compensate for students who do not take part in the testing. This is calculated independently for each sampled classroom. The basic sampling weight attached to each student record is the product of the three intermediate weights: the first-stage (school) weight, the second-stage (classroom) weight, and the third-stage (student) weight.

*Overall sampling weight.* The overall student sampling weight is the product of the three weights, including the nonparticipation adjustments.

*Scaling.* The primary approach to reporting PIRLS achievement data is based on IRT scaling methods. The IRT analysis provides a common scale on which performance can be compared across countries. Student reading achievement is summarized using a family of IRT models. In 2006 PIRLS, 2- and 3-parameter logistic IRT models were used for dichotomously scored items, and generalized partial credit models for

constructed-response items with two or three available score points. The IRT methodology is preferred for developing comparable estimates of performance for all students, since students respond to different passages and items depending upon which of the test booklets they receive. This methodology produces a score by averaging the responses of each student to the items that he or she takes in a way that takes into account the difficulty and discriminating power of each item. The approach followed in PIRLS uses information from the background questionnaires to provide improved estimates of student performance (a process known as conditioning) and multiple imputation to generate student scores (or “plausible values”) for analysis and reporting.

In addition to providing a basis for estimating mean achievement, scale scores permit estimates of how students within jurisdictions vary and provide information on percentiles of performance. Treating all participating jurisdictions equally, the PIRLS scale average across jurisdictions was set to 500 and the standard deviation to 100. Since the jurisdictions vary in size, each jurisdiction is weighted to contribute equally to the mean and standard deviation of the scale. The average and standard deviation of the scale scores are arbitrary and do not affect scale interpretation.

In the PIRLS 2001 analysis, achievement scales were produced for each of the two reading purposes—reading for literary experience and reading for information—as well as for reading overall. The PIRLS 2006 reading achievement scales were designed to provide reliable measures of student achievement common to both the 2001 and 2006 assessments, based on the metric established originally in 2001. In 2006 PIRLS, in addition to the scale for reading achievement overall, IRT scales were created to measure changes in achievement in the two purposes of reading and two overarching reading processes.

*Imputation.* No imputations are generated for missing values. However, multiple imputations are used to generate student scores (or “plausible values”) for analysis and reporting.

The PIRLS item pool is far too extensive to be administered in its entirety to any one student, and so a matrix-sampling test design was developed whereby each student is given a single test booklet containing only a part of the entire assessment. The results for all of the booklets are then aggregated using IRT techniques to provide results for the entire assessment. Since each student responds to a subset of the assessment items, multiple imputations (the generation of “plausible values”) are used to derive reliable

estimates of student performance on the assessment as a whole. Since every student proficiency estimate incorporates some uncertainty, PIRLS follows the customary procedure of generating five estimates for each student and using the variability among them as a measure of this imputation uncertainty, or error. In the PIRLS international reports (Mullis et al. 2003, 2007), the imputation error for each variable is combined with the sampling error for that variable to provide a standard error incorporating both.

## 5. DATA QUALITY AND COMPARABILITY

A group of distinguished international reading scholars, the Reading Development Group, was formed to construct the PIRLS framework and endorse the final reading assessment. Each jurisdiction followed internationally prescribed procedures to ensure valid translations and representative samples of students. The national QCMs compared the final version of the booklets with the international translation verifier's comments to ensure that their suggestions had been incorporated appropriately into the materials. The QCMs were then appointed in each jurisdiction to monitor the testing sessions at the schools to ensure that the high standards of the PIRLS data collection process were met.

### Sampling Error

The standard errors of the reading proficiency statistics reported by PIRLS include both sampling and imputation variance components.

When, as in PIRLS, the sampling design involves multistage cluster sampling, there are several options for estimating sampling errors that avoid the assumption of simple random sampling. The jackknife repeated replication technique (JRR) is chosen by PIRLS because it is computationally straightforward and provides approximately unbiased estimates of the sampling errors of means, totals, and percentages.

The particular application of the JRR technique used in PIRLS is termed a paired selection model because it assumes that the primary sampling units (PSUs) can be paired in a manner consistent with the sample design, with each pair regarded as members of a pseudo-stratum for variance estimation purposes. When used in this way, the JRR technique appropriately accounts for the combined effect of the between- and within-PSU contributions to the sampling variance. The general use of JRR entails systematically assigning pairs of schools to sampling zones, and randomly selecting one of these schools to have its contribution doubled and the other

to have its contribution zeroed, so as to construct a number of "pseudo-replicates" of the original sample. The statistic of interest is computed once for the original sample, and once again for each pseudo-replicate sample. The variation between the estimates for each of the replicate samples and the original sample estimate is the jackknife estimate of the sampling error of the statistic.

To apply the JRR technique used in PIRLS 2001 and PIRLS 2006, the sampled schools were paired and assigned to a series of groups known as "sampling zones." In total, 75 zones were used, allowing for 150 schools per jurisdiction. When more than 75 zones were constructed, they were collapsed to keep the total number to 75. For more information on sampling error, see the [PIRLS technical reports](#) (Martin, Mullis, and Kennedy 2003, 2007).

**Imputation error.** For each of the PIRLS reading scales, reading overall, and literary and informational reading, the IRT scaling procedure yields five imputed scores or plausible values for every student. The difference between the five values reflects the degree of uncertainty in the imputation process.

The general procedure for estimating the imputation variance using plausible values is the following. First compute the statistic ( $t$ ) for each set of plausible values ( $M$ ). The statistic  $t_m$ , where  $m = 1, 2, \dots, 5$ , can be anything estimable from the data, such as a mean, the difference between means, percentiles, and so forth. Once the statistics are computed, the imputation variance is then computed as

$$Var_{imp} = (1 + 1/M)Var(t_m)$$

where  $M$  is the number of plausible values used in the calculation, and  $Var(t_m)$  is the variance of the estimates computed using each plausible value.

### Nonsampling Error

Due to the particular situations of individual PIRLS jurisdictions, sampling and coverage practices have to be adaptable in order to ensure an internationally comparable population. As a result, nonsampling errors in PIRLS can be related both to coverage error and nonresponse.

**Coverage error.** PIRLS expects all participating jurisdictions to define their national desired population to correspond as closely as possible to its definition of the international desired population. Although jurisdictions are expected to include all students in the target grade in their definition of the population,

sometimes they have to reduce their coverage. Although jurisdictions were expected to do everything possible to maximize coverage of the population by the sampling plan, schools could be excluded if they were in geographically remote regions, if they were of extremely small size, if they offered a curriculum or a school structure that was different from that found in the mainstream education system, or if they provided instruction only to students in the categories defined as “within-school exclusions.”

**Table 18. Weighted U.S. response rates for 2001 and 2006 PIRLS assessments**

Year	School response rate	Student response rate	Overall response rate
2001	86	96	83
2006	86	95	82

NOTE: All weighted response rates 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. Response rates are after replacement.

SOURCE: Martin, M.O., Mullis, I.V.S., and Kennedy, A.M. (Eds.). (2003). *PIRLS 2001 Technical Report*. Boston College, International Study Center, Chestnut Hill, MA. Baer, J., Baldi, S., Ayotte, K., and Green, P. (2007). *The Reading Literacy of U.S. Fourth-Grade Students in an International Context: Results From the 2001 and 2006 Progress in International Reading and Literacy Study (PIRLS)* (NCES 2008-017). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.

Within-school exclusions were limited to students who, because of some disability, were unable to take the PIRLS tests, including educable mentally disabled students, functionally disabled students, and non-native-language speakers.

#### **Nonresponse error.**

*Unit nonresponse.* Unit nonresponse error results from nonparticipation of schools and students. Weighted and unweighted school and student response rates for PIRLS are computed for each participating jurisdiction. To monitor school participation, three school participation rates are computed: one using only originally sampled schools; one using sampled and first replacement schools; and one using sampled and both first and second replacement schools. Student participation rates are also computed, as are overall participation rates.

The minimum acceptable school-level response rate, before the use of replacement schools, was set at 85 percent. Likewise, the minimum acceptable student-

level response rate was set at 85 percent. Jurisdictions understood that the goal for sampling participation was 100 percent of all sampled schools and students. Guidelines for reporting achievement data for jurisdictions securing less than full participation were modeled after IEA’s Trends in International Mathematics and Science Study (TIMSS). Jurisdictions were assigned to one of three categories on the basis of their sampling participation. Jurisdictions in Category 1 were considered to have met the PIRLS sampling requirements and to have an acceptable participation rate. Jurisdictions in Category 2 met the sampling requirements only after including replacement schools. Jurisdictions that failed to meet the participation requirements, even with the use of replacement schools, were assigned to Category 3.

In 2001, almost all jurisdictions met the PIRLS sampling requirements and belonged in Category 1. Because they met the sampling requirements only after including replacement schools, England, the Netherlands, and the United States belonged in Category 2. Although Morocco and Scotland had overall weighted participation rates of 69 and 74 percent, respectively (even after including replacement schools), it was decided that these rates did not warrant the placement of the jurisdictions in Category 3. Instead, the results for Morocco and Scotland were annotated to indicate that they nearly satisfied the guidelines for sample participation rates after including replacement schools.

In 2006, almost all jurisdictions met the PIRLS sampling requirements and belonged in Category 1. Because they met the sampling requirements only after including replacement schools, Scotland, the United States, the Netherlands, and Belgium (Flemish) were placed in Category 2. Although Norway had overall participation rates after including replacement schools of just below 75 percent (71 percent), it was decided during the sampling adjudication that this rate did not warrant placement in Category 3. Instead, the results for Norway were annotated in the 2006 international report similarly to what was done for Morocco and Scotland in 2001.

#### **Data Comparability**

*IEA Reading Literacy Study and PIRLS.* In 1991, the IEA launched the Reading Literacy Study, which assessed the reading literacy of 4<sup>th</sup>- and 9<sup>th</sup>-grade students in 32 jurisdictions. In 2001, IEA launched PIRLS in 35 jurisdictions. Although built on the foundation of the 1991 study, PIRLS is a new and different study, with a new assessment framework describing the interaction between two major reading purposes (literary and informative) and a range of four

comprehension processes, an innovative reading test, and newly developed questionnaires for parents, students, teachers, and school principals.

Because the PIRLS 2001 reading test differed in a number of respects from the 1991 test, it was not possible to link the results of the two studies directly together. However, since PIRLS 2001 was scheduled to collect data on fourth-grade students 10 years after the 1991 Reading Literacy Study, PIRLS jurisdictions that participated in 1991 were given the opportunity of measuring changes in reading literacy achievement over that period by re-administering the 1991 reading literacy test to primary and elementary school students as part of the PIRLS data collection. The resulting study is known as the Trends in IEA's Reading Literacy Study. In 2001, nine jurisdictions replicated the 1991 Reading Literacy Study: Greece, Hungary, Iceland, Italy, New Zealand, Singapore, Slovenia, Sweden, and the United States. Conducted at the third or fourth grades (the grade with the most 9-year-olds), the study assessed student reading in three major domains: narrative texts, expository texts, and documents. Students completed a brief questionnaire about their home and school literacy activities and instruction. For more information on the trend study, see [Trends in Children's Reading Literacy Achievement 1991–2001: IEA's Repeat in Nine Countries of the 1991 Reading Literacy Study](#) (Martin et al. 2003). No such trend study was administered in conjunction with the 2006 PIRLS.

The United States conducted a study to compare the two international studies in the aspects of reading literacy each assessed, the types of texts they used in the assessments, and the types and difficulty of the questions they used. Both differences and similarities were found. The definitions of reading literacy were very similar. The types of passages used were similar, but in actually choosing and categorizing passages, the Reading Literacy Study emphasized the types of texts, while PIRLS focused on purposes for reading. In most cases, the passages and texts in PIRLS were longer, more engaging, and more complex. The question taxonomies that were generated to frame the tasks in the assessments were very different. The Reading Literacy Study taxonomy had a text focus with activities such as verbatim responses, main theme, and locating information. The PIRLS taxonomy suggested more consideration of the readers' interaction with the passage, especially in the categories of "interpret and integrate ideas and information" and "examine and evaluate content, language, and textual elements." The use of a high number of constructed-response items permitted the PIRLS questions to tap a wider range of reading responses; this is supported by the limited

analysis of a sample of questions using Wixso's Levels of Depth of Knowledge. In general, PIRLS called for a wider range of skills than did the Reading Literacy Study, especially skills requiring deeper thinking. Also, the PIRLS passages were presented in an engaging and authentic manner that might have improved students' motivation to read and respond to the texts. This is one area where the form of PIRLS might have contributed to students' level of performance. However, if students lacked the skills necessary to respond to the items, engaging texts would not have helped much. For more information on the comparison study, see the [PIRLS-IEA Reading Literacy Framework: Comparative Analysis of the 1991 IEA Reading Study and the Progress in International Reading Literacy Study](#) (Kapinus 2003).

**National Assessment of Educational Progress (NAEP) and PIRLS.** To date, there have been two studies undertaken to compare the frameworks, reading passages, and assessment items of NAEP and PIRLS. The first study compared NAEP 2002 and PIRLS 2001 at both the framework and item levels. The second study updates with analysis of the passages and item sets added in NAEP 2007 and PIRLS 2006.

**Definitions and organizations.** In terms of how the domain is defined, there is considerable overlap between the NAEP and PIRLS concepts of reading literacy. The 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).

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. 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. 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 suited below and above the targeted grade level to capture the range of reading ability.

Each of these passages has items associated with it—approximately 12-13 per passage in PIRLS and 10 per passage in NAEP. 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, although there is a greater emphasis on this in NAEP, with 69 percent of items classified as such 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 on the NAEP 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.

For more information on the similarities and differences between PIRLS and NAEP, see [A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments](#) (Binkley and Kelly 2003), and [Comparing PIRLS and PISA with NAEP in Reading, Mathematics, and Science](#) (Stephens, and Coleman, 2007).

## 6. CONTACT INFORMATION

For content information about PIRLS, contact:

Stephen Provasnik  
Phone: (202) 502-7480  
E-mail: [stephen.provasnik@ed.gov](mailto:stephen.provasnik@ed.gov)

**Mailing Address:**

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

College, International Study Center. Chestnut Hill, MA.

Gonzalez, E.J., and Kennedy, A.M. (2003). *PIRLS 2001 User Guide for the International Database*. Boston College, International Study Center. Chestnut Hill, MA.

**7. METHODOLOGY AND EVALUATION REPORTS**

Most of the technical documentation for PIRLS is published by the International Study Center at Boston College. The U.S. Department of Education, National Center for Education Statistics, is the source of several additional references listed below; these publications are indicated by an NCES number.

**General**

Kennedy, A.M., Mullis, I.V.S., Martin, M.O., and Trong, K.L. (2007). *PIRLS 2006 Encyclopedia: A Guide to Reading Education in the Forty PIRLS 2006 Countries*. Boston College, International Study Center. Chestnut Hill, MA.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., and Kennedy, A.M. (2003). *Trends in Children's Reading Literacy Achievement 1991–2001: IEA's Repeat in Nine Countries of the 1991 Reading Literacy Study*. Boston College, International Study Center. Chestnut Hill, MA.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., and Kennedy, A.M. (Eds.). (2003). *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools in 35 Countries*. Boston College, International Study Center. Chestnut Hill, MA.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., and Flaherty, C.L. (2002). *PIRLS 2001 Encyclopedia: A Reference Guide to Reading Education in the Countries Participating in IEA's Progress in International Reading Literacy Study (PIRLS)*. Boston College, International Study Center. Chestnut Hill, MA.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., and Foy, P. (2007). *PIRLS 2006 International Report: IEA's Progress in International Reading Literacy Study in Primary Schools in 40 Countries*. Boston College, International Study Center. Chestnut Hill, MA.

**Uses of Data**

Foy, P., and Kennedy, A.M. (2008). *PIRLS 2006 User Guide for the International Database*. Boston

**Survey Design**

Campbell, J.R., Kelly, D.L., Mullis, I.V.S., Martin, M.O., and Sainsburg, M. (2001). *Framework and Specifications for PIRLS Assessment 2001—2nd Edition*. Boston College, International Study Center. Chestnut Hill, MA.

Martin, M.O., Mullis, I.V.S., and Kennedy, A.M. (Eds.). (2003). *PIRLS 2001 Technical Report*. Boston College, International Study Center. Chestnut Hill, MA.

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

Mullis, I.V.S., Kennedy, A.M., Martin, M.O., and Sainsburg, M. (2006). *PIRLS 2006 Assessment Frameworks, 2nd Edition*. Boston College, International Study Center. Chestnut Hill, MA.

Ogle, L.T., Sen, A., Pahlke, E., Jocelyn, L., Kastberg, D., Roey, S., and Williams, T. (2003). *International Comparisons in Fourth-Grade Reading Literacy: Findings From the Progress in International Reading Literacy Study of 2001 (NCES 2003-073)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

**Data Quality and Comparability**

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

Binkley, M., and Kelly, D.L. (2003). *A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments (NCES Working Paper 2003-10)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

- Kapinus, B. (2003). *PIRLS-IEA Reading Literacy Framework: Comparative Analysis of the 1991 IEA Reading Study and the Progress in International Reading Literacy Study* (NCES Working Paper 2003-05). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Piessens, A., and Rust, K. (2003). *U.S. 2001 U.S. PIRLS Nonresponse Bias Analysis* (NCES 2003-21). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Stephens, M., and Coleman, M. (2007). [Comparing PIRLS and PISA with NAEP in Reading, Mathematics, and Science](#) (Working Paper). U.S. Department of Education. Washington, DC: National Center for Education Statistics.