Progress in International Reading Literacy Study (PIRLS)

Website: https://nces.ed.gov/surveys/pirls/
Updated: November 2020

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 education system (previously referred to as *education systems*). 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 *education systems*. 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 administered in 2001 to students in 35 education systems, in 2006 to students in 45 education systems, in 2011 to students in 53 education systems, and in 2016 to students in 61 education systems.

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 pertaining to fourth-grade students' reading literacy evaluation. 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 education systems; (2) describe on one international scale the literacy profiles of fourth-graders in school in each of the participating education systems; (3) describe the reading habits of fourth-graders in each participating education system; and (4) identify the home, school, and societal factors associated with the literacy levels and reading habits of fourth-graders in school.

Components
PIRLS assesses four broad-based comprehension processes within each of the two purposes for reading: focus on and retrieve explicitly stated information; make straightforward inferences, interpret and integrate ideas and information; and evaluate and critique content and textual elements.

Also, PIRLS focuses on three aspects of reading literacy: purposes for reading; processes of comprehension; and student reading behaviors and attitudes. The first two aspects are measured through the PIRLS *assessment* component, which is administered to each participating student. The third dimension, reading behaviors and attitudes, is measured through a separate component of *background questionnaires*. In 2016, the PIRLS administration included the PIRLS assessment as well as ePIRLS, an assessment of online informational reading.

PROGRESS IN INTERNATIONAL READING LITERACY STUDY:

Three assessments:

➢ PIRLS

➢ ePIRLS

➢ digitalPIRLS

Three aspects of reading literacy:

➢ Purpose for reading

➢ Processes of comprehension

➢ Reading behaviors and attitudes

Six sets of questionnaires:

➢ Student questionnaire

➢ ePIRLS student questionnaire

➢ Learning to read (home) survey

➢ Teacher questionnaire

➢ School principal questionnaire

➢ Curriculum Questionnaire

PIRLS 2021 will present a new digital web-based delivery system called digitalPIRLS. digitalPIRLS will be offered so countries can take advantage of a full computer-based assessment. The digitalPIRLS assessments will include the ePIRLS assessment of online reading initiated in 2016. With digitalPIRLS, countries will experience greater operational efficiency in translation and translation verification, data entry, and scoring, without the need for printing or shipping. digitalPIRLS will be offered as a web-based system via school-based or IEA web servers, or a USB drive connected locally to a PC with the Windows Operating System. As an alternative to digitalPIRLS, countries may administer PIRLS 2021 in paper format. ePIRLS is available in 2021 only in conjunction with digitalPIRLS. For more information on digitalPIRLS, please visit the IEA website at https://www.iea.nl/.

Source versions of all instruments (assessment booklets, the ePIRLS assessment, questionnaires, and manuals) were prepared in English and translated into the primary language or languages of instruction in each education system. In addition, it was sometimes necessary to adapt the instrument for cultural purposes, even in countries that use English as the primary language of instruction. All adaptations were reviewed and approved by the International Study Center to ensure they did not change the substance or intent of the question or answer choices.

The first aspect of the *assessment* component that is targeted by PIRLS is purposes of reading. The purposes of reading component encompasses the two main reasons why young students read printed materials: for literary experience and for the acquisition and use of information. To measure the ability of students to read for literary experience, fictional texts are used; to measure students' skills for acquiring and using information, nonfictional texts are used. In 2016, literary experience and acquiring and using information each made up 50 percent of this aspect of the PIRLS reading assessment. The second aspect of the PIRLS *assessment* component is processes of comprehension, which describes how young readers interpret and make sense of text. In 2016, this aspect was composed of four categories: focusing on and retrieving explicitly stated information (20 percent), making straightforward inferences (30 percent), interpreting and integrating ideas and information (30 percent) and evaluating and critiquing content and textual elements (20 percent).

Both PIRLS and PIRLS Literacy devote half of the assessment passages to each of the purposes for reading, while the ePIRLS online assessment focuses solely on reading to acquire and use information. The ePIRLS approach simulates websites from the Internet, through which students can navigate to accomplish school-based research projects or tasks. Because PIRLS Literacy is designed for students earlier in the process of learning to read, a larger percentage of items (50 percent of the assessment) is devoted to measuring foundational reading comprehension processes—the ability to focus on and retrieve explicitly stated information. Also, PIRLS Literacy has shorter reading passages with easier vocabulary and syntax.

The second component, *background questionnaires*, collects information on reading behaviors and attitudes (the third aspect of reading literacy targeted by PIRLS), and helps to provide a context for the performance scores. These questionnaires focus on such topics as students' attitudes and beliefs about learning, their habits and homework, and their lives both in and outside of school; teachers' attitudes and beliefs about teaching and learning, teaching assignments, class size and organization, instructional practices, and participation in professional development activities; and principals' viewpoints on policy and budget responsibilities, curriculum and instruction issues, and student behavior, as well as descriptions of the organization of schools and courses.

Assessment. In the main PIRLS assessment, each student completes a reading comprehension assessment booklet, which contains two blocks of passages that form the foundation of the PIRLS reading literacy test. Half of the assessment blocks contain passages of literary text which include realistic stories and traditional tales, and half contain passages of informational texts including chronological and non-chronological articles, a biographical article, and an informational leaflet. The passages of text are followed by questions about the text, which the student answers using constructed-response and multiple-choice response options.

The PIRLS assessment passages are submitted and reviewed by PIRLS countries and are geared to a fourth-grade level. To provide good coverage of each reading literacy domain, a total of eight blocks of test items were developed (for the 2001 assessment), representing over five hours of testing time. Each block of assessment material contained from 11 to 14 items that together represent at least 15 score points. The PIRLS reading test includes 98 items altogether, representing a total of 133 score points.

Student testing time was kept to 80 minutes per student: 40 minutes for each of the two blocks, plus an additional 15-30 minutes for the student questionnaire. Using a matrix sampling technique, the eight blocks of reading comprehension passages were used to create ten different student booklets that each contained two blocks of text. Possible booklet combinations included 1) two blocks of literary passages; 2) two blocks of informational text; or 3) one block of literary text and one block of informational text. The distribution of blocks across booklets "links" the booklets to enable the achievement data to be scaled using

item response theory methods, and enables PIRLS to report results from a representative sample for more assessment items than can fit in one booklet without making the assessment longer for the student.

The PIRLS 2006 assessment was based on ten passages: five literary and five informational. Each passage was accompanied by approximately twelve questions, with the total assessment having 126 items. Two of the literary passages and two of the informational passages (so four blocks) were retained from the 2001 assessment in order to measure trends, with the rest of the passages (three literary and three informational) being new. These ten blocks or passages were distributed across thirteen booklets for PIRLS 2006, and included 126 items across the 10 assessment blocks comprising a total of 167 score points.

The PIRLS 2011 assessment contained ten reading passages: two from 2001 and 2006, four from 2006 only, and four new passages. The assessment consisted of 135 items, representing over six hours of testing time. The ten blocks were distributed across twelve assessment booklets for 2011. The use of the six common passages (or "trend blocks") from the 2001 and 2006 assessments allows for the analysis of changes in reading literacy over the ten-year period for the education systems that participated in both assessment administrations.

In 2016, both PIRLS and PIRLS Literacy consisted of 12 passages/blocks, each of which was expected to require 40 minutes of student testing time. Six of the ten PIRLS blocks were included in previous PIRLS assessments: two in all three assessments (2001, 2006, and 2011), two in both PIRLS 2006 and PIRLS 2011, and two in PIRLS 2011 only. Ten blocks were developed specifically for PIRLS Literacy; four of the passage and item blocks were previously used in 2011 as part of prePIRLS, and two blocks came from the main PIRLS assessment.

The complete ePIRLS assessment consists of five school-based online reading tasks, each of which involves 2–3 different websites totaling 5 to 10 web pages, together with a series of comprehension questions based on the task. Similar to the PIRLS and PIRLS Literacy passages, each task with accompanying questions takes 40 minutes to complete. In order to keep student response burden to a reasonable level, each individual student completes just two ePIRLS tasks, followed by 5 minutes for a short online questionnaire.

PIRLS 2021 will include 18 passages with various response items (18 booklets for both digital format or paper format). Each booklet consists of two parts: one literary experience passage and one informational reading passage/items. PIRLS 2021 will employ a group-adaptive design. All countries will administer the same reading passages and items, but the rate at which different test forms are distributed in a country will be tailored to the population. This innovative adaptive design will improve PIRLS' measurement of reading at all levels of the distribution for countries with varying reading proficiency while also increasing student engagement. For more information on PIRLS 2021, please visit the IEA website at https://www.iea.nl/.

Background Questionnaires. An important part of the PIRLS design is a set of questionnaires targeting factors related to reading literacy (i.e., reading behaviors and attitudes). To accomplish this, questionnaires are completed by principals, teachers, parents, and students. Prior to each new PIRLS data collection, previous versions of the questionnaires are reviewed extensively by the participating countries' National Research Coordinators (NRCs), as well as by the Questionnaire Development Group, known as Questionnaire Item Review Committee (QIRC). The QIRC comprises 10–12 experienced NRCs from different participating countries who have analyzed PIRLS data and are using the data in their countries. Like the assessment items, all questionnaire items were field tested and revised if necessary prior to their inclusion in the final questionnaires.

PIRLS included four sets of background questionnaires in 2001: one each for the tested student, the parent or primary care giver of the tested student, the student's reading teacher, and the principal of the school to which the tested student attended. PIRLS 2006 also administered questionnaires to students, parents, teachers, and school principals. In addition, the NRC in each country completed a new, online curriculum questionnaire that provided data on the country's goals for reading instruction. For 2011, the background questionnaires for teachers, students, and principals were revised somewhat from previous years, and online versions of the school and teacher questionnaires were offered to respondents as the primary mode of data collection. The curriculum questionnaire was also administered. In 2016, the questionnaires were revised from 2011, and the ePIRLS student questionnaire was added.

Student questionnaire. Each student taking the PIRLS reading assessment completes the student questionnaire. The questionnaire asks about students' home environment, such as languages spoken at home, books in the home, and other home resources for learning. This questionnaire also gathers information on student experiences in school, including feelings of school belonging and whether they are victims of bullying. Finally, the student questionnaire gathers data on out-of-school reading habits and attitudes toward reading, including whether they like reading, their confidence in reading, and their engagement in reading lessons. The student questionnaire requires 15–30 minutes to complete.

ePIRLS student questionnaire. Students also participating in ePIRLS complete a brief questionnaire as part of this computer-based assessment. The questionnaire asks students about their level of competency and experience using computers and finding information on the Internet. This questionnaire requires 5 minutes to complete.

Teacher questionnaire. The reading teacher of each fourth-grade class sampled for PIRLS also completes a questionnaire. This questionnaire asks teachers about classroom contexts for reading instruction, such as characteristics of the class, reading instructional time, and instructional approaches. The questionnaire also asks about teacher characteristics, such as their career satisfaction, education, and recent professional development activities. This questionnaire requires about 35 minutes to complete.

School questionnaire. The principal of each school that is sampled is asked to respond to the school questionnaire. The principals are asked about school characteristics, such as student demographics, the school environment, and the availability of school resources and technology. The questionnaire also includes items focusing on the principal's leadership role, education, and experience. It is designed to take about 30 minutes.

Curriculum questionnaire. First used in PIRLS 2006, this questionnaire is administered to NRCs. They are asked to provide information about their national policies on reading curricula, goals and standards for reading instruction, and time specified for reading instruction, as well as information on preprimary education and teacher education policies.

Home questionnaire. The home questionnaire, entitled the Learning to Read Survey, is addressed to the parents or primary caregivers of each student taking part in the PIRLS 2016 data collection. This short questionnaire solicits information on the home context, such as languages spoken in the home, parents' reading activities and attitudes toward reading, and parents' education and occupation. The questionnaire also collects data on the students' educational activities and experiences outside of school including early childhood education, early literacy and numeracy activities, and the child's reading readiness at the beginning of primary school. This questionnaire is designed to take 10–15 minutes to complete. The home questionnaire is not administered in the United States.

Periodicity

PIRLS is administered once every 5 years, near the end of the school year in each education system. PIRLS was conducted in 2001, 2006, 2011, and 2016. The next administration is scheduled for 2021.

Data Availability

Information on the availability of data for PIRLS can be found at https://nces.ed.gov/surveys/pirls/datafiles.asp.

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 education system compare with students in another education system?

- Do fourth-grade students value and enjoy reading?

- Internationally, how do the reading habits and attitudes of students vary?

To assist in this process, the *PIRLS Encyclopedia* is created which provides a profile of each country's education system, with a particular focus on reading education for primary-school children. The encyclopedia provides general data on economic and educational indicators, describes how the education system is organized, and describes how decisions are made about education. The reading curriculum, including goals, materials, and instruction, is also discussed, along with information on the assessment of reading achievement.

3. KEY CONCEPTS

International desired population. This is the grade or age level that each education system 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, 2011, and 2016 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 education systems, the target grade was the fourth grade or its national equivalent.

National desired population. PIRLS expects all participating education systems 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 education system, then students enrolled in fourth grade were the national desired population for that education system. For PIRLS 2006, 2011, and 2016 if the fourth grade of primary school was the grade that represents 4 years of schooling in a particular education system (counting from the 1st year of

ISCED Level 1), then students enrolled in fourth grade were the national desired population for that education system.

National defined population. The national defined population is the population of students who were actually included in an education system's survey population. Although education systems are expected to include all students in the target grade in their definition of the population, sometimes it is not possible to include all students who fall under the definition of the international desired population. All students in the desired population who are not included in the defined population are referred to as the excluded population.

National Research Coordinators (NRCs). Each country appoints a National Research Coordinator who, together with staff at the PIRLS national center, is responsible for all aspects of the study within that country. NRCs play a central part in ensuring the suitability of the assessment materials including scoring and coding of data, and they are responsible for collecting and preparing data for the PIRLS assessment according to the procedures specified internationally, and even more specifically is responsible for implementing the sample design, including documenting each step of the sampling procedure. NRCs also nominate quality control monitors and interact with data processing centers to accomplish data cleaning and documentation.

Reading literacy. PIRLS joins the terms reading and literacy to convey a broad notion of what the ability to read means—a notion that includes the ability to reflect on what is read and to use it as a tool for attaining individual and societal goals. The term "reading literacy" has been used by IEA since naming its 1991 Reading Literacy Study, and it remains the appropriate term for what is meant by "reading" and what PIRLS is assessing. In developing a definition of reading literacy to serve as the basis for PIRLS, the Reading Development Group for 2001 looked to IEA's 1991 study, in which reading literacy was defined as "the ability to understand and use those written language forms required by society and/or valued by the individual." The Reading Development Group for 2001 elaborated on this definition for PIRLS so that it applies across ages yet makes explicit reference to aspects of the reading experience of young children. Beginning with PIRLS 2006, the definition was refined to highlight the widespread importance of reading in school and everyday life: "Reading literacy is defined as the ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment". This view of reading reflects numerous theories of reading literacy as a constructive and interactive process.

4. SURVEY DESIGN

Target Population
Fourth-grade student population. The target 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. This target grade was usually the fourth grade of primary school. Because fourth grade generally signals the completion of formal reading instruction, countries for which the target grade would have been the third grade were permitted to retain the fourth grade as their target grade. The PIRLS 2001 target population was derived from that used by TIMSS in 1995 and was identical to that used by TIMSS 2003 at the primary school level.

For PIRLS 2006, the target population was defined as all students enrolled in the fourth grade of formal schooling, counting from the first year of primary school as defined by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) International Standard Classification for Education (ISCED). Accordingly, the fourth year of formal schooling was the fourth grade in most countries.

The target population for PIRLS 2011 was all students enrolled in the grade that represents four years of schooling, counting from the first year of ISCED Level 1. ISCED provides an international standard for describing levels of schooling across countries. The ISCED system describes the full range of schooling, from preprimary (Level 0) to the second level of tertiary education (Level 6). ISCED Level 1 corresponds to primary education or the first stage of basic education. Four years later would be the PIRLS target grade, which is the fourth grade in most countries. However, given the linguistic and cognitive demands of reading, PIRLS is designed to avoid assessing very young children. Thus, countries were recommended to assess the next higher grade (i.e., fifth grade) if the average age of fourth grade students at the time of testing was less than 9.5 years.

For most countries participating in PIRLS 2011, the target grade was fourth grade. However, in England, Malta, New Zealand, and Trinidad and Tobago, children begin primary school at an early age. Therefore, these countries assessed students in the fifth year of schooling and their students were still among the youngest in PIRLS 2011.

Several new initiatives were introduced in 2011 that affected the target population in several countries. One new initiative was prePIRLS, which was developed as a less difficult version of PIRLS to provide more assessment options for developing countries where students may not be prepared for the demands of PIRLS. prePIRLS was based on the same view of reading comprehension as PIRLS but was designed to assess basic reading skills that were a

prerequisite for success on PIRLS. Botswana, Colombia, and South Africa administered prePIRLS to their fourth grade students. Colombia also administered PIRLS to the same fourth grade students, providing a basis for a link between the PIRLS and prePIRLS scales. As well, in 2011, PIRLS was given to students in the fifth or sixth grades in countries where the assessment might be too difficult for their fourth grade students. Accordingly, Botswana, Honduras, Kuwait, and Morocco chose to administer PIRLS in both the sixth and fourth grades.

The target population for PIRLS 2016 was the same as for 2011. The PIRLS 2016 cycle also included PIRLS Literacy—a new, less difficult reading literacy assessment, and ePIRLS—an extension of PIRLS with a focus on online informational reading.

Teacher population. The target teacher population consists of all teachers linked to the selected students. Note that these teachers are therefore not a representative sample of teachers within an education system. Rather, they are the teachers who teach a representative sample of students in grade 4 within the education system.

School population. The target school population consists of all eligible schools containing one or more fourth-grade classrooms.

Sample Design

PIRLS uses a two-stage stratified cluster sample design. The first stage consists of a sample of schools, which may be stratified; the second stage consists of a sample of one or more classrooms from the target grade in sampled schools.

First-stage sampling selects individual schools with a probability proportionate to size (PPS) approach, which means that the probability is proportional to the estimated number of students enrolled in the target grade. Substitution schools are also selected to replace any schools that are originally sampled but refuse to participate. The original and substitution schools are selected simultaneously. In the second stage of sampling, one or two fourth-grade classes are randomly sampled in each school.

PIRLS guidelines call for a minimum of 150 schools to be sampled in each education system, with a minimum of 4,000 students assessed. A sample of 150 schools yields 95 percent confidence limits for school-level and classroom-level mean estimates that are precise to within 16 percent of their standard deviations. Countries with small class sizes or less than 30 students per school are directed to consider sampling more schools, more classrooms per school, or both, to meet the minimum target of 4,000 tested students. For countries choosing to participate in both PIRLS and PIRLS Literacy, the required student sample size is doubled—i.e., around 8,000 sampled students. Countries could choose to select more schools or more classes within

sampled schools to achieve the required sample size. Because ePIRLS is designed to be administered to students also taking PIRLS, the PIRLS sample size requirement remains the same for countries choosing also to participate in ePIRLS.

In the United States, the PIRLS 2001 sample consisted of 3,763 fourth-grade students from 174 schools (after substitution). In 2006, the U.S. sample consisted of 5,190 fourth-grade students from 183 schools (after substitution).

For the 2011 data collection, there were 370 U.S. schools, after substitution, consisting of 12,726 fourth-grade students. The reason for a larger sample in 2011 was due to the coinciding administration of the Trends in International Mathematics and Science Study (TIMSS). To accommodate this concurrent administration, schools with at least two grade 4 classrooms were asked to participate in both studies, with one classroom being randomly assigned to TIMSS and the other to PIRLS.

In the United States, one sample was drawn to represent the nation at grade 4 for PIRLS 2011. In addition to this national sample, a state public school sample was also drawn at grade 4 for Florida, which chose to participate in PIRLS separately from the nation in order to benchmark their student performance internationally. The sample frame for public schools in the United States was based on the 2011 National Assessment of Educational Progress (NAEP) sampling frame. The 2011 NAEP sampling frame was based on the 2007–08 Common Core of Data (CCD). The PIRLS 2011 data for private schools were from the 2007–08 Private School Universe Survey (PSS). Any school containing at least one grade 4 class was included in the school sampling frame.

The U.S. PIRLS 2016 national school sample consisted of 176 schools, which was higher than the international sampling minimum of 150 to offset anticipated school nonresponse and ineligibility. A total of 158 U.S. schools agreed to participate in PIRLS 2016, including 131 from the original sample and 27 sampled as replacements for nonparticipating schools from the original sample. Of the 158 U.S. schools that participated in PIRLS, 153 also participated in ePIRLS. In total, 4,425 U.S. students participated in PIRLS and 4,090 of these students also participated in ePIRLS.

The U.S. sampling frame was explicitly stratified by three categorical variables: Poverty status (high or low, defined by percentage of students eligible for free or reduced-price lunch); type of school (public or private); and region of the country (Northeast, Central, West, Southeast). The U.S. sample was implicitly stratified (that is, sorted for sampling) by two categorical variables: locality (four levels) and minority status (above or below 15 percent of the student population).

Implementation

PIRLS is sponsored by the IEA and carried out under a contract with the TIMSS & PIRLS International Study Center and the data collection contractor. The National Center for Education Statistics, in the Institute of Education Sciences at the U.S. Department of Education, is responsible for the implementation of PIRLS in the United States. PIRLS emphasizes the use of standardized procedures in all participating education systems, so that each education system collected its own data based on comprehensive manuals and training materials. These materials explain the survey's implementation, including precise instructions for the work of school coordinators and scripts for test administrators to use in testing sessions. The International Study Center monitors compliance with the standardized procedures.

The PIRLS 2001 instruments were translated into 35 languages. The PIRLS 2006 instruments were again prepared in English and then translated into 45 languages. Although most countries administer the assessment in just one language, there have been some exceptions. For example, in 2006, nine countries plus the five Canadian provinces administered PIRLS in two languages, Spain administered the assessment in its five official languages, and South Africa administered the assessment in eleven languages. To ensure comparability among translated instruments, the International Study Center established guidelines and reviewed and approved all adaptions. For PIRLS 2011, the assessment was translated into 45 different languages.

The PIRLS 2016 assessment instruments were translated into 40 different languages, across 50 participating countries and 6 benchmarking entities, the PIRLS Literacy assessment instruments were translated into 10 languages across 6 countries, and the ePIRLS assessment instruments were translated into 14 languages across 14 countries and 2 benchmarking entities. Of these participants, 24 countries and 4 benchmarking entities administered the instruments in more than one language.

Data Collection and Processing

The IEA provides overall support in coordinating PIRLS. The Secretariat, located in Amsterdam, has particular responsibility for membership, translation verification, and hiring the quality control monitors. The Data Processing and Research Center, located in Hamburg, is responsible for the accuracy and consistency of the PIRLS database within and across countries.

Reference dates. PIRLS is administered near the end of the school year in each education system. For PIRLS 2001, in education systems 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 Southern

Hemisphere where the school year typically ends in November or December, the assessment was conducted in October or November 2001.

For PIRLS 2006, education systems 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 assessment was conducted in October 2005.

For PIRLS 2011, the education systems in the Southern Hemisphere conducted the study between October and December 2010. Education systems in the Northern Hemisphere conducted the assessment between March and June 2011.

For PIRLS 2016, the education systems in the Southern Hemisphere conducted the study between October and December 2015. Education systems in the Northern Hemisphere conducted the assessment between March and June 2016.

Data collection and cleaning. Each country was responsible for carrying out all aspects of the data collection by using standardized procedures developed for the study. Manuals provided 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.

The International Study Center monitored compliance with the standardized procedures. NRCs were asked to nominate one or more persons unconnected with their national center, such as retired school teachers, to serve as quality control monitors for their education systems. The International Study Center developed manuals for the monitors and briefed them in 2-day training sessions about PIRLS, the responsibilities of the national centers in conducting the study, and their own roles and responsibilities. For the 2001 PIRLS test administration, 15 schools in each country were observed. For 2006, ten percent of the schools' test administrations were visited by monitors, and for PIRLS 2011, some 30 of the 370 schools in the sample were visited by monitors. For PIRLS 2016, International Quality Control Monitors observed 814 PIRLS/PIRLS Literacy testing sessions and 209 ePIRLS testing sessions.

The NRC in each education system was responsible for the scoring and coding of data in that education system, following established guidelines. The NRC and, sometimes, additional staff attended scoring training sessions held by the International Study Center. The training sessions focused on the scoring rubrics and coding system employed in PIRLS. Participants in these training sessions were provided extensive practice in scoring example items over several days. Information on within-education-system

agreement among coders was collected and documented by the International Study Center. Information on scoring and coding reliability was also used to calculate cross-education-system agreement among coders.

The NRC from each education system was responsible for data entry. In the United States, the data collection contractor collected data for PIRLS 2016 and entered the data into data files with a pre-specified, common international format. IEA-supplied data-entry software (WinDEM) facilitated the checking and correction of data by providing various data consistency checks. The data were then sent to the IEA Data Processing Center (DPC) in Hamburg, Germany, for cleaning. The DPC checked that the international data structure was followed; checked the identification system within and between files; corrected single case problems manually; and applied standard cleaning procedures to questionnaire files. Results of the data cleaning process were documented by the DPC. This documentation was then sent to the NRC along with any remaining questions about the data. The NRC then provided the DPC with revisions to coding or solutions for anomalies. The DPC subsequently compiled background univariate statistics and preliminary test scores based on classical item analysis and item response theory (IRT).

Estimation Methods
Before the collected data are analyzed, student records are assigned sampling weights to ensure that student representation in the PIRLS analysis closely matches the prevalence of groups in the student population for the grade assessed. Under the PIRLS sample design, schools and students have unequal but known probabilities of selection; as a consequence, file-supplied sampling weights must be applied to analysis and subsequent results, in order to generalize to the population.

After sample weights are assigned, scaling and estimation can be conducted. During the scaling phase, IRT procedures are used to estimate the measurement characteristics of each assessment question. During the estimation phase, the results of the scaling are used to produce estimates of student achievement. Subsequent analyses relate the achievement results to the background variables collected by PIRLS.

Weighting. Students are assigned sampling weights to adjust for over- or under-representation of particular groups in the final sample. When students are weighted, none of the data are discarded and each student contributes to the results for the total number of students represented. The weight assigned to a student is therefore the inverse of the probability that the student was selected for the sample. The use of sampling weights is necessary for the computation of sound, nationally representative estimates. Weighting also adjusts for various situations such as school and student

nonresponse because data cannot be assumed to be randomly missing. All PIRLS 2001, 2006, 2011, and 2016 analyses are conducted using sampling weights and 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 regions. A school level participation adjustment is then made in the school weight to compensate for any sampled schools that did not participate and were not replaced. That adjustment is calculated independently for each explicit stratum.

Classroom weight. In the second step, a classroom weight reflecting the probability of the sampled classroom(s) being selected from among all the classrooms in the school at the target grade level is calculated. This weight is calculated independently for each participating school. If a sampled classroom in a school did not participate, or if the participation rate among students in a classroom fell below 50 percent, a classroom-level participation adjustment is made to the classroom weight. Classroom participation adjustment could occur only within "participating schools" (a school was considered as a "participating school" if and only if there was at least one sampled classroom with at least 50 percent of its students participating in the study). If one of at least two selected classrooms in a school did not participate, the classroom participation adjustment is computed at the explicit stratum level, rather than at the school level, to reduce the risk of bias.

Student weight. The third and final step consists of calculating a student weight. For most PIRLS participants, intact classrooms are sampled, so each student in the sampled classrooms is certain of selection, making the student weight 1.0. When students are further sampled within classrooms, a student weight reflecting the probability of the sampled students being selected within the classroom is calculated. A nonparticipation adjustment is then made to adjust for sampled students who did not take part in the testing. This adjustment is calculated independently for each sampled classroom.

Overall (basic) sampling weight. The overall student sampling weight is the product of the three weights just described and includes any nonparticipation adjustments that were made.

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. 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 item responses of each student, taking into account the difficulty and discriminating ability of each item. To enable comparisons across PIRLS assessments, common test items are included in successive administrations, and any item parameters that change dramatically are treated as unique items.

The propensity of students to answer questions correctly is estimated for PIRLS using a two-parameter IRT model for dichotomous constructed response items, a three-parameter IRT model for multiple choice response items, and a generalized partial credit IRT model for polytomous constructed-response items. The scale scores assigned to each student were estimated using a plausible values procedure, with input from the IRT results. With IRT, the difficulty of each item, or item category, is deduced using information about how likely it is for students to get some items correct (or to get a higher rating on a constructed response item) versus other items. Once the parameters of each item are determined, the ability of each student can be estimated even when different students have been administered different items. At this point in the estimation process achievement scores are expressed in a standardized logit scale. In order to make the scores more meaningful and to facilitate their interpretation, the scores for the PIRLS 2001 assessment are transformed to a scale with a mean of 500 and a standard deviation of 100.

To make PIRLS 2006 scores comparable to 2001 scores, the 2001 and 2006 data for countries that participated in both years were first scaled together, to estimate item parameters. Ability estimates for all students in the 2001 and 2006 assessment were then estimated based on the new item parameters. A linear transformation was then applied to put these estimates on the 2001 metric so that the jointly calibrated 2001 scores have the same mean and standard deviation as the original 2001 scores. This also preserves any differences in average scores between the 2001 and 2006 waves of assessment.

To make PIRLS 2011 scores comparable to 2001, these steps are repeated for each pair of 2006 and 2011 data: two adjacent years of data are jointly scaled, then resulting ability estimates are linearly transformed so that the mean and standard deviation of the prior year is preserved. As a result, the transformed 2011 scores are comparable to all previous waves of assessment and longitudinal comparisons between all waves of data are meaningful.

To provide results for the PIRLS 2016 assessment on the PIRLS achievement scales, the 2016 proficiency scores (plausible values) for overall reading had to be transformed to the PIRLS reporting metric. This was accomplished through a set of linear transformations as part of the concurrent calibration approach. The linear transformation constants were obtained by first computing the international means and standard deviations of the proficiency scores for the overall reading scale using the plausible values produced in 2011 based on the 2011 item calibrations for the trend countries. These were the plausible values published in 2011. Next, the same calculations were done using the plausible values from the re-scaled PIRLS 2011 assessment data based on the 2016 concurrent item calibration for the same set of countries. There are five sets of transformation constants for the PIRLS reading scale, one for each plausible value. The trend countries contributed equally in the calculation of these transformation constants. These linear transformation constants were applied to the overall reading proficiency scores and for all participating countries and benchmarking participants. This provided student achievement scores for the PIRLS 2016 assessment that are directly comparable to the scores from all previous assessments.

Much like the normal PIRLS scaling procedure, the PIRLS Literacy scaling approach involved the same four tasks of calibrating the achievement items, creating principal components for conditioning, generating proficiency scores, and placing these proficiency scores on the PIRLS reading reporting scale.

The ePIRLS scaling methodology adopted the same four steps of calibration, conditioning, generating proficiency scores, and placing those scores on the PIRLS reading scale.

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.

Plausible values, estimation, multiple imputation. Most cognitive skills testing is concerned with accurately assessing the performance of individual respondents, for the purposes of diagnosis, selection, or placement. Regardless of the measurement model used—whether classical test theory or item response theory—the accuracy of these measurements can be improved (i.e., the amount of measurement error can be reduced) by increasing the number of items given to the individual. Thus, it is common to see achievement tests designed to provide information on individual students that contain more than 70 items. For the distribution of proficiencies in large populations, however, more efficient estimates can be obtained from a matrix sampling design like that used in PIRLS. This design solicits relatively few responses from each sampled respondent while maintaining a wide range of content

representation when responses are aggregated across all respondents. With this approach, however, the advantage of estimating population characteristics is offset by the inability to make precise statements about individuals. The uncertainty associated with individual estimates becomes too large to be ignored, and aggregations of individual student scores can lead to seriously biased estimates of population characteristics.

Plausible values methodology is a way to address this issue by using all available data to estimate directly the characteristics of student populations and subpopulations and then to generate multiple imputed scores (plausible values) from these distributions, which can be used in analyses with standard statistical software. For PIRLS, plausible values are estimated to characterize students participating in the assessment, given their background characteristics.

As mentioned, plausible values are imputed values and are not test scores for individuals in the usual sense. In fact, they are biased estimates of the proficiencies of individual students. Plausible values do, however, provide unbiased estimates of population characteristics (e.g., means and variances of demographic subgroups), and represent what the performance of an individual on the entire assessment might have been, had it been observed. Plausible values are estimated as random draws (usually five) from an empirically derived distribution of score values based on the student's observed responses to assessment items and on background variables. Each random draw from the distribution is considered a representative value from the distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses. Differences between plausible values drawn for a single individual quantify the degree of error (the width of the spread) in the underlying distribution of possible scale scores that could have caused the observed performances.

Recent Changes
There have been several important changes to the PIRLS assessment since 2001.

- PIRLS 2001 pioneered the Learning to Read Survey, completed by students' parents or caregivers, as well as the PIRLS Encyclopedia, comprised of chapters written by each participating country describing its reading curriculum and instruction.

- In 2006, PIRLS was expanded to report results by comprehension processes in addition to literary and informational reading purposes.

- In 2006, greater emphasis was given to the PIRLS Curriculum Questionnaire completed by each participating country.

- In 2011, the PIRLS and TIMSS assessment cycles came together, providing a unique opportunity for countries to collect reading, mathematics, and science achievement data on the same fourth grade students.

- Also in 2011, IEA introduced a less difficult version of PIRLS—called prePIRLS—as a way for countries with developing education systems to assess reading at the end of primary school and as a stepping stone to participating in PIRLS.

- In 2016, PIRLS was expanded to include two new assessments of reading comprehension, PIRLS Literacy and ePIRLS. The PIRLS Literacy assessment is equivalent to PIRLS in scope and reflects the same conception of reading as PIRLS, except it is less difficult overall. ePIRLS is an innovative assessment of online reading, designed to be responsive to the information age and provide important information about how well students are developing 21st century skills.

Future Plans
The next administration is scheduled for 2021. PIRLS 2021 will be the new assessment, digitalPIRLS, which includes PIRLS and ePIRLS, and will be administered in a complete computer-based delivery system.

5. DATA QUALITY AND COMPARABILITY

Comparisons made in PIRLS (e.g. education systems' averages compared to the U. S. average) are tested for differences using statistical significance, which requires the estimation of standard errors. However, the estimation of correct standard errors is complicated by the complex sample and assessment designs of PIRLS: both the sample design and assessment design generate error variance and mandate a set of statistically complex procedures. For PIRLS, estimates produced using the data are subject to two types of error—nonsampling and sampling error. Nonsampling error can be due to errors made in collecting and processing data. Sampling error can occur because the data were collected from a sample rather than a complete census of the population.

Sampling Error
Sampling errors arise when a sample of the population, rather than the whole population, is used to estimate a statistic. Different samples from the same population would likely produce somewhat different estimates of the statistic in question. This means that there is a degree of uncertainty associated with statistics estimated from a sample. This

uncertainty, or sampling variance, is usually expressed as the standard error of a statistic estimated from sample data. For PIRLS, there is the additional complexity of the multi-stage cluster and assessment matrix sampling designs, which result in estimated standard errors containing both a sampling variance component—estimated by a jackknife repeated replication (JRR) procedure—and an additional imputation variance component arising from the assessment design.

The matrix sampling design assigns a single test assessment booklet containing only a portion of the PIRLS assessment to each individual student. Using the scaling techniques described above, results are aggregated across all booklets to provide results for the entire assessment, with plausible values being generated as estimates of student performance on the assessment as a whole. The variability among these are combined with the sampling error for that variable, to provide a standard error that incorporates both error components. The correctly estimated standard errors are then used to conduct *t*-tests that compare other education system averages to the U.S. average, for example, and to construct confidence intervals.

Confidence intervals provide a way to make inferences about population statistics in a manner that reflects the sampling error associated with the statistic. Assuming a normal distribution, the population value of this statistic can be inferred to lie within a 5-percent confidence interval in 95 out of 100 replications of the measurement on different samples drawn from the same population. For example, the average reading score for U.S. fourth-grade students was 549 in 2016, and this statistic had a standard error of 3.1. Therefore, it can be stated with 95 percent confidence that the actual average of U.S. fourth-grade students in 2016 was between 543 and 555.

Nonsampling Error

Nonsampling error is a term used to describe variations in the estimates that may be caused by population coverage limitations, nonresponse bias, and measurement error, as well as data collection, processing, and reporting procedures. The sources of nonsampling error are typically problems like unit and item nonresponse, the difference in respondents' interpretations of the meaning of the survey questions, response differences related to the particular time the survey was conducted, and mistakes in data preparation.

One strategy implemented by PIRLS to reduce nonresponse bias is the a priori identification of replacement schools. Ideally, response rates to study samples should always be 100 percent, and although the PIRLS administrators worked hard to achieve this goal, it was anticipated that a 100 percent participation rate would not be possible in all countries. To avoid sample size losses, the PIRLS sampling plan identified, a priori, replacement 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. Replacement 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. Although the use of replacement schools did not eliminate the risk of nonresponse bias, employing implicit stratification and ordering the school sampling frame by size increased the chances that any sampled school's replacements would have similar characteristics. This approach maintains the desired sample size while restricting replacement schools to strata where nonresponse occurred.

IEA-developed participation or response rate standards are next applied. These standards were set using composites of response rates at the school, classroom, and student and teacher levels, and response rates were calculated with and without the inclusion of the replacement/substitute schools. These standards took the following two forms for 2016: Category 1-education system met the standards, having 85 percent minimum school and student participation rates and 95 percent classroom participation rates; and Category 2-education system met the standards after substitution. Countries satisfying the category 1 standard are included in the international tabular presentations without annotation. Those able to satisfy only the category 2 standard are included as well but are annotated to indicate their response rate status. The data from education systems failing to meet either standard (identified as Category 3 in previous PIRLS administrations) are presented separately in the international tabular presentations. Table PIRLS-1 displays response rates for the U.S. for the 2001, 2006, 2011, and 2016 administrations of PIRLS and ePIRLS.

Data Comparability

From its inception, PIRLS was designed to measure trends in reading literacy achievement. Many of the countries participating in PIRLS 2016 also participated in the previous study cycles in 2001, 2006, and 2011. As a result, these countries have the opportunity to measure progress in reading achievement across four time points: 2001, 2006, 2011, and 2016. In order to ensure comparability of the data across participating education systems, the IEA provides detailed international requirements on the various aspects of data collection, and implements quality control procedures. Participating countries are obliged to follow these requirements, which pertain to target populations, sampling design, sample size, exclusions, and defining participation rates.

In the United States, data used by NCES on fourth-grade students' reading achievement comes primarily from two sources: NAEP and PIRLS. There are distinctive differences between PIRLS and NAEP. A comparative

study was conducted of PIRLS 2011 and NAEP 2009/2011, which overall suggested that the NAEP 2011 reading assessment may be more cognitively challenging than PIRLS 2011 for U.S. fourth-grade students and that caution should be exercised when attempting to compare fourth-grade students' performance on PIRLS 2011 with fourth-grade students' performance on the NAEP 2011 reading assessment.

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).

Table PIRLS-1. Weighted U.S. response rates for PIRLS assessments: 2001, 2006, 2011, and 2016

Year	School response rate	Student response rate	Overall response rate
2001	86	96	83
2006	86	95	82
2011	85	96	81
2016 main assessment	92	94	86
2016 ePIRLS	89	90	80

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: PIRLS methodology reports; available at https://nces.ed.gov/pubsearch/getpubcats.asp?sid=099.

6. CONTACT INFORMATION

For content information on PIRLS, contact:

Sheila Thompson
Phone: (202) 245-8330
E-mail: Sheila.Thompson@ed.gov

Mailing Address
National Center for Education Statistics
Institute of Education Sciences
Potomac Center Plaza
550 12th Street, SW
Washington, DC 20202

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.

https://timssandpirls.bc.edu/pirls2006/encyclopedia.html.

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. https://timss.bc.edu/pirls2001i/PIRLS2001_Pubs_ER.html.

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. https://timssandpirls.bc.edu/pirls2006/intl_rpt.html.

Mullis, I.V.S., Martin, M.O., Minnich, C.A., Drucker, K.T., & Ragan, M.A. (Eds.). (2012). *PIRLS 2011 Encyclopedia: Education Policy and Curriculum in Reading, Volumes 1 and 2*. Boston College, International Study Center. Chestnut Hill, MA. https://timssandpirls.bc.edu/pirls2011/encyclopedia-pirls.html.

Uses of Data
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.

Herget, D., Dalton, B., Kinney, S., Smith, W.Z., Wilson, D., and Rogers, J. (2019). *Progress in International Reading Literacy Study (PIRLS) and ePIRLS 2016: U.S. Public-Use Data Files* (NCES 2019-114). U.S. Department of Education. Washington, DC: National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2019114

Herget, D., Dalton, B., Kinney, S., Smith, W.Z., Wilson, D., and Rogers, J. (2019). *Progress in International Reading Literacy Study (PIRLS) and ePIRLS 2016: U.S. Restricted-use Datafile* (NCES 2019-115). U.S. Department of Education. Washington, DC: National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2019115

Mullis, I.V.S., Martin, M.O., F o y , P ., and Drucker, K . T ., (Eds.). (2012). *PIRLS 2011 International Results in Reading*. Boston College, International Study Center. Chestnut Hill, MA & International Association for the Evaluation of Educational Achievement IEA Secretariat: Amsterdam, the Netherlands. https://timssandpirls.bc.edu/pirls2011/downloads/P11_IR_FullBook.pdf.

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. https://timss.bc.edu/pirls2001i/PIRLS2001_Pubs_IR.html

Survey Design

Herget, D., Dalton, B., Kinney, S., Smith, W.Z., Wilson, D., and Rogers, J. (2019). *U.S. PIRLS and ePIRLS 2016 Technical Report and User's Guide* (NCES 2019-113). U.S. Department of Education. Washington, DC: National Center for Education Statistics. https://nces.ed.gov/pubs2019/2019113.pdf.

Martin, M.O., and Mullis, I.V.S. (Eds.) (2015). *PIRLS 2016 Assessment Framework, 2nd Edition*. Boston College, International Study Center. Chestnut Hill, MA. https://timssandpirls.bc.edu/pirls2016/downloads/P16_Framework_2ndEd.pdf.

Martin, M. O., Mullis, I. V. S., & Hooper, M. (Eds.). (2017). *Methods and Procedures in PIRLS 2016.* Boston College, International Study Center. Chestnut Hill, MA.

https://timssandpirls.bc.edu/publications/pirls/2016-methods.html.

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. https://timss.bc.edu/pirls2001i/PIRLS2001_Pubs_TR.html.

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. https://timssandpirls.bc.edu/pirls2006/tech_rpt.html.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M, Trong, K. L., and Sainsbury, M. (2012). *PIRLS 2011 Assessment Framework*. Boston College, International Study Center. Chestnut Hill, MA. https://timssandpirls.bc.edu/pirls2011/downloads/PIRLS2011_Framework.pdf.

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 Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubs2003/2003073.pdf.

Thompson, S., Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., and Jenkins, F. (2012). *Highlights from PIRLS 2011: Reading Achievement of U.S. Fourth-Grade Students in an International Context* (NCES 2013–010). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Government Printing Office. https://nces.ed.gov/pubs2013/2013010.pdf.

Warner-Griffin, C., Liu, H., Tadler, C., Herget, D., and Dalton, B. (2017). *Reading Achievement of U.S. Fourth-Grade Students in an International Context: First Look at the Progress in International Reading Literacy Study (PIRLS) 2016 and ePIRLS 2016* (NCES 2018-017). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. https://nces.ed.gov/pubs2018/2018017.pdf.

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. https://nces.ed.gov/pubs2008/2008017_1.pdf.

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. https://nces.ed.gov/pubs2003/200310.pdf.

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. https://timss.bc.edu/pirls2001i/pirls2001_pubs_trr.html.

U.S. Department of Education, National Center for Education Statistics*. PIRLS-IEA Reading Literacy Framework: Comparative Analysis of the 1991 Reading Study and Progress in International Reading Literacy Study, NCES 2003–05*, by Barbara Kapinus. Project Officer: Laurence Ogle. Washington, DC: 2003. https://nces.ed.gov/pubs2003/200305.pdf.

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. https://nces.ed.gov/Surveys/PISA/pdf/comppaper12082004.pdf.