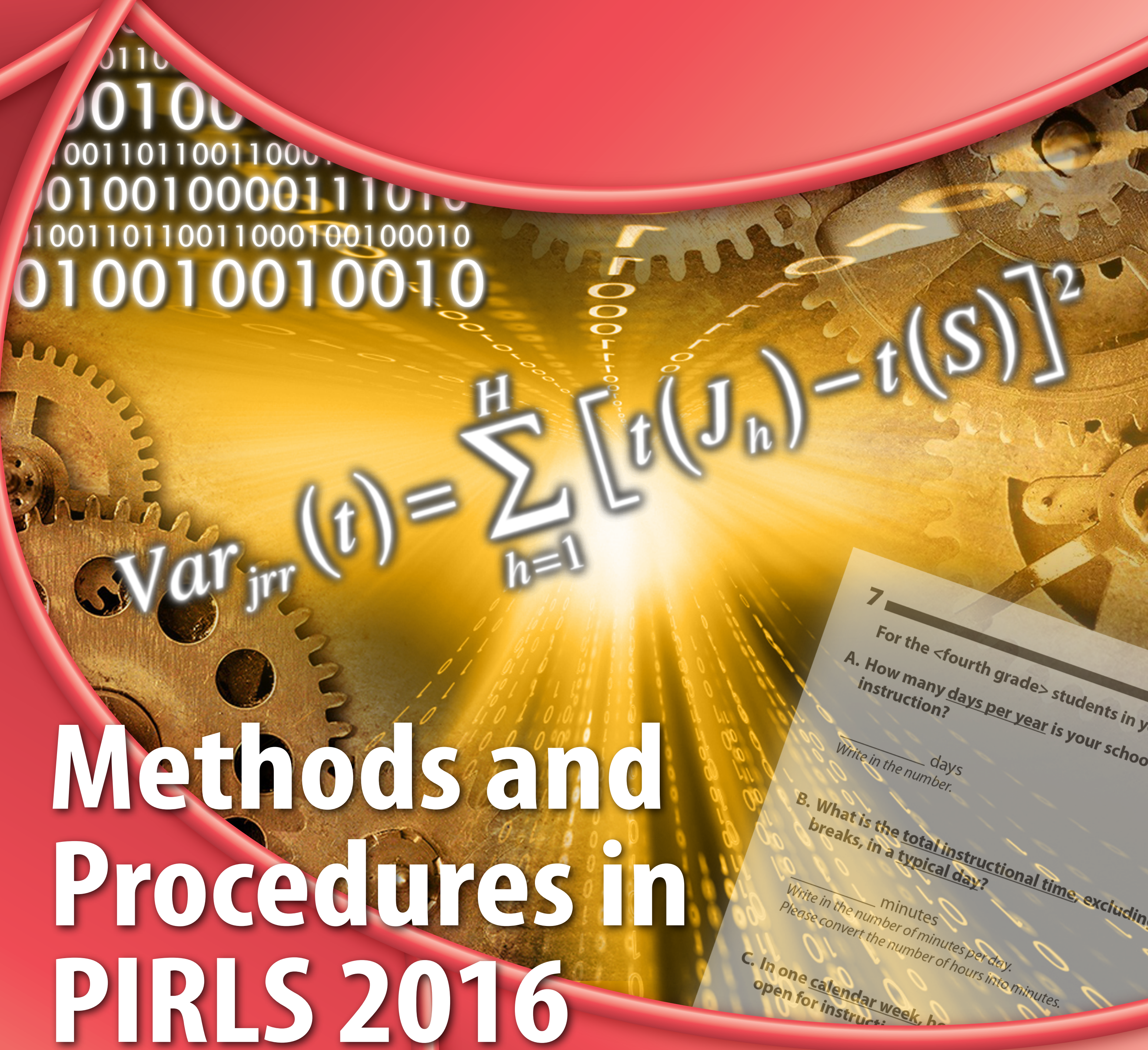
PROGRESS IN INTERNATIONAL READING LITERACY STUDY

PIRLS



Methods and Procedures in PIRLS 2016

Michael O. Martin,
Ina V.S. Mullis,
Martin Hooper,
Editors

IEA
TIMSS & PIRLS
International Study Center
Lynch School of Education
BOSTON COLLEGE

Copyright © 2017 International Association for the Evaluation of Educational Achievement (IEA)
METHODS AND PROCEDURES IN PIRLS 2016
Edited by: Michael O. Martin, Ina V.S. Mullis, and Martin Hooper

Publishers: TIMSS & PIRLS International Study Center,
Lynch School of Education, Boston College
and
International Association for the Evaluation of Educational Achievement (IEA)

Library of Congress Catalog Card Number: 2017951161
ISBN: 978-1-889938-44-8

For more information about PIRLS contact:
TIMSS & PIRLS International Study Center
Lynch School of Education
Boston College
Chestnut Hill, MA 02467
United States

tel: +1-617-552-1600
fax: +1-617-552-1203
e-mail: pirls@bc.edu
pirls.bc.edu

Boston College is an equal opportunity, affirmative action employer.

INSTRUMENT DEVELOPMENT

SAMPLING

DATA COLLECTION PROCEDURES

PIRLS 2016

Methods and Procedures in PIRLS 2016

Instrument Development

CHAPTER 1

Developing the PIRLS 2016 Achievement Items

Ina V.S. Mullis
Caroline O. Prendergast

Unique Characteristics of the 2016 PIRLS Assessment

The general approach to developing the PIRLS achievement items is similar from assessment cycle to assessment cycle, but each assessment cycle tends to have some unique characteristics that influence instrument development. Besides providing measures on another cycle for the PIRLS trend lines monitoring changes in educational achievement, 2016 also was remarkable for two reasons.

- It was the inaugural year of the ePIRLS extension of PIRLS. ePIRLS was introduced in 2016 to assess online informational reading skills in a simulated Internet environment and was administered via computer (PCs). In ePIRLS, students are assessed on their ability to acquire and use information from webpages while investigating science and social studies topics through authentic, school-like assignments.
- The PIRLS Reading Achievement scale was extended to include PIRLS Literacy, which took the place of prePIRLS. PIRLS Literacy 2016 advanced prePIRLS by linking a less difficult version of the PIRLS assessment to the well-established PIRLS metric to enable assessing reading comprehension across a broader range of countries. Countries whose students were not yet prepared to take PIRLS were still able to participate in this important international project by administering PIRLS Literacy. Countries' results for the two different versions are both reported on the PIRLS scale.

ePIRLS 2016: Extending PIRLS to Assess Online Reading

Recognizing that the Internet has become the primary source for obtaining information at work, at home, and for school, PIRLS 2016 was extended to include ePIRLS on a voluntary basis for countries already participating in PIRLS and where students were familiar with using computers. ePIRLS used an engaging simulated Internet environment to measure fourth grade students' achievement in reading for informational purposes. The assessment was administered via computer using a PC platform. Countries were responsible for using their own computers. In most cases, countries used the computers available in the schools or arranged for rental computers. ePIRLS presented students with authentic school-like assignments about science and social studies topics, which align with purposes for school reading. Led by a teacher avatar, students were asked to navigate through multiple, interconnected webpages containing both textual and visual information. ePIRLS allows for assessing reading comprehension skills beyond those used in “traditional” print material.

In addition to the data collected through the PIRLS 2016 Context Questionnaires (see Chapter 2), ePIRLS has its own short student questionnaire pertaining to students’ familiarity with computers and online reading. Also, some process data will be analyzed to study students’ navigation patterns.

PIRLS Literacy 2016

For a variety of reasons, there are some countries where most children in the fourth grade are still developing fundamental reading skills. Therefore, IEA offers options for matching the PIRLS reading assessment to the country’s educational development. For some countries, the PIRLS Literacy version of PIRLS is a better match with students’ learning. New for 2016, PIRLS Literacy has been placed on the same scale as PIRLS, with the two versions (PIRLS and PIRLS Literacy) having four passages in common with one another.

PIRLS Literacy reflects the same conception of reading as does PIRLS except the assessment is less difficult. The PIRLS Literacy assessment is consistent with the PIRLS framework for assessing reading comprehension. However, typically the passages are shorter with less complex syntax, and the questions include a different mix of items across the comprehension processes compared to PIRLS. PIRLS Literacy places somewhat greater emphasis on straightforward retrieval of information compared to PIRLS and less emphasis on straightforward inferencing, interpreting and integrating ideas and information, and evaluating and critiquing content and textual elements.

PIRLS Literacy was developed together with PIRLS. It uses the same context questionnaires, and the expert committees reviewed both the PIRLS and PIRLS Literacy passages, items, and scoring guides together. The challenge was identifying a range of passages with content suitable for fourth grade students that could be used in PIRLS, in both PIRLS and PIRLS Literacy, and only in PIRLS Literacy. Also, PIRLS Literacy passages contain questions placed throughout the passages to enable students to answer questions as they proceed through the text, rather than the PIRLS approach of presenting the entire passage followed by the set of questions.

The PIRLS Approach to Measuring Trends

Because PIRLS is designed to measure trends, the assessments cannot change dramatically from cycle to cycle. That is, PIRLS is based on a well-known premise for designing trend assessments (ascribed to John Tukey and Albert Beaton):

> "If you want to measure change, do not change the measure."

However, the achievement tests and questionnaires also need to be updated with each cycle to prevent the assessments from becoming dated and no longer relevant to current learning goals and policy issues. It is important that the content reflects the most recent discoveries in the field and is presented in ways consistent with students' instructional and everyday experiences.

To maintain continuity with past assessments while keeping up with current topics and technology, the PIRLS assessments evolve with each cycle. PIRLS has a specific design for rotating passages and items out of the assessment after each cycle and replacing them with newly developed passages and items for the following cycle. The remaining assessment items are kept secure to be readministered in subsequent cycles.

The design for passage/item replacement provides for each assessment to include passages and items from three cycles—essentially, one-third newly developed, one-third from the previous cycle, and one-third from two cycles before. With permission from IEA the replaced assessment passages and items are available on a restricted use basis for educational and research purposes (please see http://www.iea.nl/copyright-notice for permissions information).

Overview of the PIRLS 2016 Development Process

According to the PIRLS assessment design, it is necessary to replace a specific portion of the passages and achievement items for each upcoming cycle. Although the majority of the assessment items are carried forward from the previous assessment cycle to measure trends, the task of updating the instruments for each new cycle—every five years for PIRLS since 2001—is a substantial undertaking. All of the passages, and subsequently the items, must be reviewed by experts and agreed upon by the diverse participating countries.

The TIMSS & PIRLS International Study Center at Boston College uses a collaborative process to select the passages and develop the new items needed for each PIRLS cycle. A broad overview of the process includes:

- Updating the frameworks for the upcoming assessment
- Identifying and selecting appropriate reading passages
- Developing items and their scoring guides in accordance with the frameworks

- Conducting a full-scale field test
- Selecting the new assessment items based on the frameworks, field test results, and to complement existing passages and items from previous cycles
- Conducting training in how to reliably score responses to constructed response items (i.e., questions to which students provide a written response rather than choosing from a set of options)

The development process is directed and managed by the staff of the TIMSS & PIRLS International Study Center at Boston College, who collectively have considerable experience in the measurement and assessment of reading achievement. For PIRLS 2016, Executive Directors Ina Mullis and Michael Martin managed the assessment development process.

Also playing a key role in achievement item development were the National Research Coordinators (NRCs) designated by their countries to be responsible for the complex tasks involved in implementing PIRLS in their countries. The TIMSS & PIRLS International Study Center worked with the NRCs and experts from the countries throughout the development process to identify suitable PIRLS passages and develop new test items. To provide additional subject-matter expertise and support, staff consulted closely with external reading specialists. Continuing from PIRLS 2006 and 2011, the PIRLS 2016 Chief Reading Consultant was Marian Sainsbury, National Foundation for Educational Research (NFER), London, England. The Reading Development Group (RDG) provided additional advice and guidance in developing the PIRLS assessment through periodic reviews. The countries participating in PIRLS nominate RDG members for each PIRLS cycle.

Exhibit 1.1 lists the eight members of the PIRLS 2016 RDG.

Exhibit 1.1: PIRLS 2016 Reading Development Group (RDG)

Julian Fraillon
Australian Council for Educational Research
Australia

Jan Mejding
Aarhus University
Department of Education
Denmark

Galina Zuckerman
Russian Academy of Education
Russian Federation

Elizabeth Pang
Ministry of Education
Singapore

Jenny Wiksten Folkeryd
Uppsala University
Sweden

Ahlam Habeeb Msaiqer
Abu Dhabi Education Council
United Arab Emirates, Abu Dhabi

Donald Leu
University of Connecticut
United States

Karen Wixson
University of North Carolina, Greensboro
United States

RDG members met four times for PIRLS 2016. At the first RDG meeting in Copenhagen, Denmark (July 2013), the RDG reviewed the reading frameworks, potential passages, ePIRLS prototypes, and draft item writing guidelines. At the second meeting in London, England (April 2014), the RDG reviewed PIRLS field test passages and items and ePIRLS field test tasks and items. At the third meeting in Stockholm, Sweden (July 2015), the RDG reviewed field test results and made recommendations to the NRCs regarding which passages and items to include in the 2016 assessments. At the final meeting in Lübeck, Germany (May 2017), the RDG conducted the PIRLS 2016 scale anchoring process (see Chapter 13).

During busy periods in between RDG meetings, the Chief Reading Consultant and several RDG members served as a task force to assist in completing specific tasks, such as updating the framework (PIRLS Framework Task Force) or developing items (PIRLS Item Development Task Force).

The PIRLS 2016 Development Schedule

To accomplish the development work in timely fashion, the assessment was developed over three years of the five-year cycle according to a specific timeline. Essentially, one year or so was devoted to updating the framework and identifying appropriate passages, the second year was devoted to item development, and the third year to conducting the field test and selecting the materials for data collection. (The fourth year of the cycle was data collection and the fifth was analysis and reporting.)

Exhibit 1.2 shows the PIRLS 2016 development schedule from updating the frameworks to data collection.

Exhibit 1.2: PIRLS 2016 Development Schedule for Achievement Items

Date(s)		Group and Activity
July-December	2012	To begin work on updates to the Assessment Framework for PIRLS 2016, the TIMSS & PIRLS International Study Center summarized the curricular emphases in reading described in the *PIRLS 2011 Encyclopedia*
December	2012	Task Force of reading experts proposed updates for the 2016 Assessment Framework, incorporating information from the Encyclopedia (Boston, USA)
January	2013	TIMSS & PIRLS International Study Center sent proposed Assessment Framework updates to National Research Coordinators (NRCs) in preparation for the 1st NRC Meeting
February	2013	TIMSS & PIRLS International Study Center presented plans for ePIRLS, including a sample task, and NRCs reviewed proposed updates to Assessment Framework at 1st NRC meeting (Hamburg, Germany)
February-July	2013	TIMSS & PIRLS International Study Center incorporated feedback from 1st NRC meeting to further refine the *PIRLS 2016 Assessment Framework*
March-September	2013	NRCs submitted and reviewed proposed reading passages in preparation for the 2nd NRC meeting (Portorož, Slovenia)
March-July	2013	TIMSS & PIRLS International Study Center developed prototype ePIRLS tasks
May	2013	NRCs received a promotional ePIRLS video, which illustrated ePIRLS using the Polar Bears task
July	2013	Reading Development Group (RDG) reviewed proposed Assessment Framework, passages, ePIRLS prototypes, and draft *PIRLS 2016 Item Writing Guidelines* at the first RDG meeting (Copenhagen, Denmark)
August	2013	TIMSS & PIRLS International Study Center updated *PIRLS 2016 Item Writing Guidelines*
September	2013	TIMSS & PIRLS International Study Center prepared final drafts of *PIRLS 2016 Assessment Framework*, incorporating RDG and NRC comments
September	2013	NRCs performed final review of the *PIRLS 2016 Assessment Framework*, selected passages, reviewed storyboards for ePIRLS tasks, and developed draft field test items at the 2nd NRC meeting (Portorož, Slovenia)
October-January	2013-2014	TIMSS & PIRLS International Study Center further refined draft field test items and scoring guides and continued to develop ePIRLS tasks
November	2013	TIMSS & PIRLS International Study Center published *PIRLS 2016 Assessment Framework (1st edition)*
February	2014	PIRLS/PIRLS Literacy Item Development Task Force reviewed and edited draft field test items and scoring guides (Boston, USA)
March-April	2014	ACER and AIR conducted cognitive labs for two sample ePIRLS tasks
April	2014	RDG reviewed PIRLS field test passages and items for PIRLS as well as storyboards for six ePIRLS tasks and items at 2nd RDG meeting (London, England)
April-May	2014	TIMSS & PIRLS International Study Center revised draft field test passages and tasks, as well as their items and scoring guides, to address RDG comments
May	2014	NRCs reviewed and approved PIRLS/PIRLS Literacy field test passages and items and reviewed storyboards for five ePIRLS tasks at 3rd NRC meeting (Dublin, Ireland)
May-July	2014	TIMSS & PIRLS International Study Center assembled field test passages and items into assessment booklets

Exhibit 1.2: PIRLS 2016 Development Schedule for Achievement Items (Continued)

Date(s)		Group and Activity
July	2014	TIMSS & PIRLS International Study Center posted PIRLS field test achievement booklets for NRCs
August	2014	ePIRLS NRCs reviewed storyboards and items for Rivers
September	2014	TIMSS & PIRLS International Study Center posted PIRLS Literacy field test achievement booklets for NRCs
October	2014	NRCS received final storyboards for Mars, Rainforests, Blackwell, Migration, Troy, and Rivers as well as the ePIRLS student questionnaire
October-December	2014	TIMSS & PIRLS International Study Center worked with each of five English-speaking countries to administer PIRLS to several classes to collect student responses to constructed response items in order to develop scoring training materials
November	2014	TIMSS & PIRLS International Study Center administered PIRLS Literacy passages in a range of classrooms in the Boston area to collect student responses to constructed response items in order to develop scoring training materials
November	2014	TIMSS & PIRLS International Study Center posted ePIRLS tasks, software, system check, online translation system, and test administrator manual for the pilot test
November	2014	ePIRLS pilot test conducted in Australia, Ireland, and Canada (Ontario) to test the ePIRLS tasks and software in a classroom setting and inform scoring guides and training materials
November-February	2014-2015	TIMSS & PIRLS International Study Center posted systems and materials for the ePIRLS field test
December	2014	PIRLS/PIRLS Literacy Item Development Task Force modified scoring guides for constructed response items based on student responses and developed scoring training materials for 4th NRC meeting (Boston, USA)
February	2015	*PIRLS 2016 Assessment Framework (2nd edition)* published online, incorporating the introduction to PIRLS Literacy and the new integrated PIRLS/PIRLS Literacy assessment design
February	2015	ePIRLS Task Force reviewed students' typed responses from the pilot and developed scoring training materials for 4th NRC meeting (Boston, USA)
February	2015	NRCs received scoring training for PIRLS, PIRLS Literacy, and ePIRLS 2016 constructed response field test items at 4th NRC meeting (Floriana, Malta)
March-April	2015	Countries conducted PIRLS, PIRLS Literacy, and ePIRLS 2016 field tests
April-May	2015	Countries submitted field test achievement data for analysis and review
June	2015	PIRLS/PIRLS Literacy Item Development Task Force reviewed field test item statistics
June	2015	ePIRLS Task Force reviewed field test item statistics
June-July	2015	TIMSS & PIRLS International Study Center assembled proposed PIRLS/PIRLS Literacy passages and items in preparation for the 3rd RDG meeting
July	2015	RDG reviewed proposed PIRLS/PIRLS Literacy passages and items in conjunction with field test results and reviewed five proposed ePIRLS tasks via computer at the 3rd RDG meeting (Stockholm, Sweden)
July	2015	ePIRLS NRCs received "Preparing Computers for ePIRLS" instructions

Exhibit 1.2: PIRLS 2016 Development Schedule for Achievement Items (Continued)

Date(s)		Group and Activity
August	2015	NRCs reviewed and approved PIRLS/PIRLS Literacy passages and items and ePIRLS storyboards for PIRLS 2016 data collection at 5th NRC meeting (Jyväskylä, Finland)
August	2015	IEA Hamburg provided information to NRCs about ePIRLS software and operations at the 5th NRC meeting (Jyväskylä, Finland)
August	2015	TIMSS & PIRLS International Study Center distributed PIRLS/PIRLS Literacy and ePIRLS 2016 data collection achievement materials to NRCs
September	2015	ePIRLS NRCs received access to the Online Translation System for main data collection
October-December	2015	Southern Hemisphere countries conducted PIRLS 2016 data collection
October	2015	TIMSS & PIRLS International Study Center updated and prepared materials for PIRLS/PIRLS Literacy 2016 constructed response scoring training
November	2015	NRCs from Southern Hemisphere countries received scoring training for PIRLS/PIRLS Literacy constructed response items (Buenos Aires, Argentina)
November	2015	TIMSS & PIRLS International Study Center finalized scoring guides and training materials for PIRLS/PIRLS Literacy and ePIRLS constructed response items and distributed them to NRCs
February-March	2016	NRCs from Northern Hemisphere countries received scoring training for PIRLS/PIRLS Literacy and ePIRLS constructed response items at 6th NRC meeting (Hong Kong SAR)
March-June	2016	Northern Hemisphere countries conducted PIRLS/PIRLS Literacy and ePIRLS 2016 data collection

Updating the Assessment Framework for PIRLS 2016

Updating the PIRLS assessment for 2016 began with reviewing and modifying the assessment framework that describes the aspects of reading comprehension to be assessed.

The basic structure of the PIRLS assessment framework is based on two dimensions: purposes for reading and processes of comprehension. Reading for literary experience and reading to acquire and use information are the two major purposes assessed by PIRLS because they account for many of the reading experiences of young children.

The four comprehension processes assessed by PIRLS are:

- Focusing on and retrieving explicitly stated information
- Making straightforward inferences
- Interpreting and integrating ideas and information
- Evaluating and critiquing content and textual elements

For PIRLS 2016, the name of the fourth comprehension process was changed to "Evaluate and Critique Content and Textual Elements" from "Examine and Evaluate Content, Language, and Textual Elements" in 2011. This newer category name better describes the processes students use when answering items assigned to this category, clarifying for item writers the kinds of items to be developed. Also, a new section was added to the framework that described the components of online reading that should be addressed in ePIRLS.

The NRCs from the participating countries discussed the framework updates at their first meeting. Following the discussion at the 1st NRC meeting in Hamburg in February 2013, the NRCs consulted with their national experts about the PIRLS updates for 2016. Next, the RDG reviewed and revised the frameworks. Using an iterative process, the NRCs once again reviewed the RDG's revised version of the framework, which was updated a final time prior to publication of the 1st edition in November 2013.

Following that, however, further discussions with the NRCs revealed dissatisfaction with the 2011 design where prePIRLS was reported separately from PIRLS. Thus, the PIRLS 2016 design was updated to strengthen the assessment of reading for children still developing fundamental reading skills. PIRLS Literacy was developed to extend the PIRLS achievement scale to address the needs of a broader range of countries. PIRLS Literacy is equivalent in scope to PIRLS, and they are linked with four passages in common. This enables results for both assessments to be reported on the same PIRLS scale. However, the new design necessitated updating the PIRLS 2016 Assessment Framework, and a 2nd edition was published in February 2015. The first chapter of the *PIRLS 2016 Assessment Framework (2nd Edition)* describes the aspects of reading comprehension to be assessed by PIRLS 2016 in detail.

Identifying Reading Passages for PIRLS and PIRLS Literacy

In total, 18 new passages and item sets needed to be developed and field tested for PIRLS and PIRLS Literacy 2016. The PIRLS 2016 assessment required field testing 12 passages (8 of which were newly developed PIRLS passages and 4 of which were newly developed to be shared between PIRLS and PIRLS Literacy), which included a total of 203 new items. In addition to the four shared passages, the PIRLS Literacy component also required field testing 6 new passages, which included a total of 173 items.

Identifying appropriate passages for the PIRLS and PIRLS Literacy assessments was critical to their success, because readers make meaning from text in a variety of ways, depending not only on the purpose for reading but also on the difficulty of the text and the reader's prior knowledge. Examples of literary texts include contemporary short stories as well as traditional tales and fables. Informational texts can be from a variety of sources, such as informational books, textbooks, and journal articles and may include graphic support in the form of charts, tables, or diagrams.

At the beginning of the assessment cycle, the TIMSS & PIRLS International Study Center sent a call for passages to all NRCs. The criteria for suitable passages was discussed at the first NRC meeting in Hamburg in February 2013. In general, the PIRLS 2016 Chief Reading Consultant, Marian Sainsbury from NFER, explained that passages should:

- Be suitable for fourth grade students in content, interest, and reading ability
- Be well written in terms of depth and complexity to allow for a sufficient number of questions
- Avoid bias in that they are sensitive to cultural differences and are likely to be equally familiar or unfamiliar to all students

In March 2013, the TIMSS & PIRLS International Study Center created a discussion board so NRCs could review passages as they were submitted. At the same time, TIMSS & PIRLS International Study Center staff and the Chief Reading Consultant also began the search for suitable materials.

In conjunction with a qualitative evaluation of each text's characteristics and appropriateness for different languages and cultures, text length and readability guided passage selection. The TIMSS & PIRLS International Study Center computed the word count and readability for each passage as a quantitative check of the grade appropriateness of the recommended texts. The Flesch-Kincaid Grade Level Formula[1] was used as a measure of readability for this purpose because of its suitability for a wide range of texts and its extensive use in education. This quantitative information was provided alongside the texts to NRCs for their review.

The NRCs and the RDG conducted an iterative passage review process at meetings and online. During the year or so allocated to find texts, the NRCs and the RDG reviewed hundreds of passages in order to identify the approximately 18-20 passages for PIRLS and PIRLS Literacy that were needed to develop items for the field test. The TIMSS & PIRLS International Study Center relied on the professional judgment of the NRCs and their within-country experts to evaluate the grade appropriateness, translatability, and cultural suitability of the texts for their students.

The TIMSS & PIRLS International Study Center prepares an international version of all the PIRLS and ePIRLS assessment items in English. Subsequently, the items are translated by participating countries into their languages of instruction with the goal of creating high quality translations that are appropriately adapted for the national context and at the same time are internationally comparable. Therefore, a significant portion of the development and review effort by NRCs is dedicated to ensuring that the passages can be translated accurately.

1 See Kincaid, Fishburne, Rogers, and Chissom (1975).

Developing Website Texts and Items for ePIRLS

Reading for informational purposes on the Internet requires many of the same reading comprehension skills and strategies as does reading offline. However, reading online also requires some new skills and strategies and is done in a different environment containing a wider variety of texts. Developing ePIRLS involved creating six tasks that included simulated Internet webpages with multiple pages of text, and included 115 items in total.

Developing appropriate and engaging webpages for each ePIRLS assessment task involved creating a variety of texts that fit into an integrated website focused on a science or social studies topic. The texts included written descriptions and explanations, diagrams, interactive images and maps, and animated graphics. ePIRLS website text development followed the same guidelines as for PIRLS passages, taking into consideration suitability for fourth grade students regarding content, interest, reading ability, complexity, and cultural sensitivity.

Especially since it was for the first time, developing the ePIRLS tasks was extremely arduous and time consuming. The TIMSS & PIRLS International Study Center developed four ePIRLS tasks, all based on the [TIMSS 2015 Science Framework] for the fourth grade. The first task developed, called "Polar Bears," was about how the melting ice in the northern Polar Regions is affecting the habitat of the polar bears. The idea of a website about polar bears was part of the presentation on extending PIRLS 2016 to assess online reading that the TIMSS & PIRLS International Study Center made at the 1st PIRLS NRC meeting, where both ePIRLS and the topic of polar bears were well received by the NRCs.

There was considerable information about the polar bears topic on the Internet including a variety of texts and images. Developing the ePIRLS task proceeded slowly, involving sorting through choices and creating simulated websites that could be examined by the students. The staff at the TIMSS & PIRLS Study Center, including the Executive Directors, the Director of the Production Department, and the Communications Specialist, carefully selected the webpages for each of several websites (e.g., about polar bears, maps and data about polar ice melting, and opinions about the future of polar bears) and drafted the narrative for the teacher avatar, Mr./Ms. Webster. The TIMSS & PIRLS International Study Center owes a debt of gratitude to Dr. Don Leu who pioneered the idea of the teacher avatar and was a member of the PIRLS 2016 RDG.

The teacher avatar guided the students through the websites in the polar bear task, asking various questions about the information in the webpages. Answering the questions required students to navigate to the appropriate webpages and read various content. For some questions, students could choose their answers from multiple-choice questions or drop-down menus, and for other questions they were asked to type in their answers.

Once the Internet images were selected, the ideas for the websites created, and the script was drafted, the production staff at the TIMSS & PIRLS International Study Center prepared storyboards covering the Polar Bear task from beginning to end. The Polar Bear storyboards provided the foundation for disseminating and reviewing the idea of ePIRLS with the NRCs and the RDG, and also allowed the TIMSS & PIRLS International Study Center to consult with Dr. Leu and his staff about how programming the tasks would work. Eventually, the Polar Bear task became the basis for a video the TIMSS & PIRLS International Study Center prepared to explain the characteristics of ePIRLS.

While the TIMSS & PIRLS International Study Center was working on task development, IEA Hamburg was working on the systems necessary to administer ePIRLS via PC. This included: an online translation system, a systems check, ePIRLS software so that USB sticks could be used to load the assessment tasks onto the countries' computers, provision to upload the student data to the IEA server in Hamburg, and a system in Hamburg to capture the data for scoring. The online translation system enables translators to adapt the international version of the ePIRLS tasks, including items and website text, into a target language directly in the online system. Additionally, the online translation system enables translators to review, revise, and verify translated text. The system check program allows test administrators to quickly check whether a given computer is able to support the ePIRLS software as delivered by the USB sticks or a local server. The data monitoring system allows NRCs to monitor collected student data through an online portal. The online scoring system streamlines the scoring process by providing scorers with student responses, scoring guides, and scoring capabilities for constructed response items.

Subsequent to the work with the Polar Bear task and the creation of the video, the TIMSS & PIRLS International Study Center developed three more tasks for the ePIRLS field test. The tasks were based on science topics and were developed using the same procedure of identifying websites and drafting a script. Then storyboards were developed, reviewed by the NRCs and RDG, and revised. Only then, were storyboards given to IEA Hamburg for programming the ePIRLS software.

Led by RDG member Dr. Julian Fraillon, the Australian Council for Educational Research (ACER) also developed three ePIRLS tasks. These tasks were in social science areas and followed a similar development path. The ideas and concepts were discussed with the TIMSS & PIRLS International Study Center and those selected for further development were then plotted out. The websites/webpages and scripts were reviewed by the TIMSS & PIRLS International Study Center before ACER drafted storyboards. The draft storyboards were thoroughly reviewed by the NRCs and RDG. The TIMSS & PIRLS International Study Center made the final revisions to the storyboards and forwarded them to IEA Hamburg for programming.

In the spring of 2014, the TIMSS & PIRLS International Study Center arranged for the American Institute for Research (AIR) to conduct cognitive labs in Washington, D.C. and ACER to conduct them in Camberwell, Victoria, Australia. Two ePIRLS tasks with 38 items in total were presented to approximately 21 students using an initial version of the test administration software. These students were observed and prompted to answer questions about the clarity, difficulty, and familiarity of the item content and format, as well as questions about the simulated Internet environment and teacher avatar. As the students completed the tasks, their interactions with the software were monitored and recorded in order to collect information about the strengths and weaknesses of the software and the testing experience. The TIMSS & PIRLS International Study Center received the cognitive lab reports in the summer of 2014.

Based on the information from the cognitive labs, six ePIRLS tasks were developed, reviewed, and programmed for inclusion in the ePIRLS pilot. The pilot took place in October and November 2014 in Australia, Ireland, and Canada (Ontario). This process provided an additional opportunity to monitor the implementation of the ePIRLS software in a classroom setting while collecting student responses to the constructed response items. The typed responses gathered during the pilot test were then used to develop scoring guides for the constructed response items for the ePIRLS field test.

Writing and Reviewing the PIRLS/PIRLS Literacy 2016 Field Test Items and Scoring Guides

The TIMSS & PIRLS International Study Center uses a collaborative process involving the participating countries to develop test items and scoring guides for the field tests. Most of the 2nd PIRLS NRC meeting in Portorož, Slovenia in September 2013 was devoted to a workshop for developing the field test items. The NRCs, together with experienced item writers from participating countries and staff from the TIMSS & PIRLS International Study Center, created the newly developed items for the PIRLS and PIRLS Literacy passages.

Prior to the PIRLS item writing workshop, TIMSS & PIRLS International Study Center staff members identified the scope of the item writing task for the field test, examining the weight given to each purpose and comprehension process in the *PIRLS 2016 Assessment Framework*, as well as how many passages and items existed from previous assessments.

In preparation for the item writing workshop, the TIMSS & PIRLS International Study Center updated the Item Writing Guidelines, an item writing manual specifically developed for PIRLS assessments. The *PIRLS 2016 Item Writing Guidelines* contain general information about procedures for obtaining good measurement (for instance, items should be independent and not provide clues to the correct responses of other items) as well as specific information on how to deal with translation issues. The manual also includes the necessary steps for developing scoring guides, as well as checklists for reviewing the PIRLS 2016 items.

At the PIRLS/PIRLS Literacy item writing workshop, country representatives were divided into teams and given specific item writing assignments to ensure that enough field test items were developed in each of the purposes and processes of comprehension areas specified in the PIRLS 2016 framework. The TIMSS & PIRLS International Study Center staff and consultants used the Item Writing Guidelines to provide training to the teams on item writing procedures for the PIRLS assessments. Once teams had completed their item writing assignments, each team reviewed the items drafted by other teams. In addition, some teams continued to send items to the TIMSS & PIRLS International Study Center for several weeks after the item writing workshop.

Exhibit 1.3 shows the number of participants in the PIRLS/PIRLS Literacy 2016 item writing workshop and the number of items written.

Exhibit 1.3: PIRLS/PIRLS Literacy 2016 Item Writing Workshop to Develop Field Test Items

Attendees	
Number of Countries and Benchmarking Entities	44
Number of Country Representatives	83
Approximate Number of Field Test Items Written at Item Writing Workshop	
PIRLS	394
PIRLS Literacy	134

Following the item writing workshop, the TIMSS & PIRLS International Study Center thoroughly reviewed the draft set of passages and field test items. Reviewers included the chief consultant and consultants experienced in developing assessment items such as those from NFER and ACER, as well as RDG members with particular item writing skills.

Finally, prior to field test instrument production, the PIRLS 2016 RDG members reviewed the proposed field test passages and items, followed by the NRCs at the 3rd NRC meeting in Dublin, Ireland in May 2014. The TIMSS & PIRLS International Study Center implemented the suggested revisions, produced the field test materials, and provided the final international version of the field test booklets to the NRCs so that they could begin translating the field test materials into their languages of instruction.

The PIRLS, PIRLS Literacy, and ePIRLS 2016 Field Tests

Because the TIMSS & PIRLS International Study Center generally field tests twice the number of passages and items actually required, the field test included the target number of new passages and items needed approximately multiplied by two. This included a total of 18 newly developed passages across PIRLS and PIRLS Literacy—8 passages for PIRLS, 4 passages to be shared in common between PIRLS and PIRLS Literacy, and 6 passages for PIRLS Literacy. Given that the field tests for PIRLS and PIRLS Literacy both included the passages in common, the PIRLS field test included 12 passages with 203 items and the PIRLS Literacy field test included 10 passages with 173 items.

The PIRLS and PIRLS Literacy field tests followed typical PIRLS procedures, where they served as full-scale “dress rehearsals” operationally for the assessments. That is, the data collection and scoring procedures to be employed in the assessments were practiced in the field test. In addition, the field tests provided important information about how well each prospective item functioned and provided a basis for selecting items for the assessments. For the countries participating in ePIRLS, the PIRLS field test students were tested again via computer, typically on the day following the PIRLS field test. The ePIRLS field test involved schools using the ePIRLS software and systems as well as the students responding to the tasks.

All materials and operational procedures for PIRLS/PIRLS Literacy 2016 and ePIRLS were field tested with samples of students selected according to rigorous sampling procedures. The field tests were designed to be conducted in approximately 30 schools in each country. This yielded approximately 9,000 student responses to each PIRLS item, approximately 1,000 for each PIRLS Literacy item, and approximately 5,000 for ePIRLS. The school samples for the PIRLS 2016 field tests and assessments were drawn simultaneously, using the same random sampling procedures. This ensures that field test samples closely approximate assessment samples, and that a school is selected for either the field test or the assessment, but not both. For example, if a country needed 150 schools for the assessment and another 30 for the field test, then a larger sample of 180 schools was selected and a systematic sample of 30 schools was selected from the 180 schools.

Because ePIRLS was a brand new computer-based online reading assessment, preparation for the ePIRLS field test was quite complicated. It involved loading the ePIRLS software onto each computer and checking the compatibility of the computer with the software. The requirement that ePIRLS students also participated in PIRLS was part of the ePIRLS field test because ePIRLS is an extension of PIRLS. The countries participating in ePIRLS field tested ePIRLS with the same students that had already participated in PIRLS, typically on the day after the PIRLS field test.

The ePIRLS field test involved 13,701 students in 13 countries and 5 benchmarking entities. Implementing and monitoring the field test involved newly developed web based systems, including the online translation system, the online scoring system, and online data monitor. The ePIRLS tasks were delivered to the students' computers via USB sticks. Responses collected during the field test were used to evaluate the measurement properties of each item. Additionally, information about students' basic navigation behavior through the hyperlinks, tabs, and advertisements in the tasks was collected in order to analyze the ways students moved through and interacted with the test administration system. The item data and the navigation data were used to revise the ePIRLS tasks and items before the main data collection.

The PIRLS, PIRLS Literacy, and ePIRLS 2016 field tests were conducted in March–April 2015. Student responses were used to evaluate the measurement properties of each field test assessment item. Exhibits 1.4 through 1.6 provide a detailed summary of the field test effort, including the number of students, teachers, and schools that participated and the number of passages and items listed by format, purpose, and comprehension process.

Exhibit 1.4: Overview of the PIRLS 2016 Field Test

	PIRLS	PIRLS Literacy	ePIRLS
Passages/Tasks	12	10	6
Total Items	**203**	**173**	**115**
Responses per item (approx.)	9,000	1,000	5,000
Participants			
Countries	49	7	13
Benchmarking Entities	7	1	5
Students	58,078	6,795	13,701
Teachers	3,025	389	–
Schools	1,634	245	561

Note that four passages and their corresponding items are common to both the PIRLS and PIRLS Literacy assessments.

Exhibit 1.5: PIRLS 2016 Number of Field Test Items by Reading Purpose and Item Format

Reading Purpose	Number of Passages/ Tasks	Number of Multiple-Choice Items	Number of Constructed Response Items	Total Number of Items	Total Number of Score Points	Percentage of Score Points
PIRLS						
Literary	6	45	56	101	130	49%
Informational	6	44	58	102	135	51%
Total	**12**	**89**	**114**	**203**	**265**	
PIRLS Literacy						
Literary	5	41	46	87	104	51%
Informational	5	44	42	86	99	49%
Total	**10**	**85**	**88**	**173**	**203**	
ePIRLS						
Informational	6	44	71	115	153	100%
Total	**6**	**44**	**71**	**115**	**153**	

Note that four passages and their corresponding items are common to both the PIRLS and PIRLS Literacy assessments.

Exhibit 1.6: **PIRLS 2016 Number of Field Test Items by Comprehension Process and Item Format**

Comprehension Process	Number of Multiple-Choice Items	Number of Constructed Response Items	Total Number of Items	Total Number of Score Points	Percentage of Score Points
PIRLS					
Focus on and Retrieve Explicitly Stated Information	30	34	64	72	27%
Make Straightforward Inferences	37	23	60	70	26%
Interpret and Integrate Ideas and Information	9	39	48	85	32%
Evaluate and Critique Content and Textual Elements	13	18	31	38	14%
Total	**89**	**114**	**203**	**265**	
PIRLS Literacy					
Focus on and Retrieve Explicitly Stated Information	33	52	85	92	45%
Make Straightforward Inferences	30	13	43	47	23%
Interpret and Integrate Ideas and Information	7	21	28	46	23%
Evaluate and Critique Content and Textual Elements	15	2	17	18	9%
Total	**85**	**88**	**173**	**203**	
ePIRLS					
Focus on and Retrieve Explicitly Stated Information	11	14	25	25	16%
Make Straightforward Inferences	15	20	35	41	27%
Interpret and Integrate Ideas and Information	6	28	34	61	40%
Evaluate and Critique Content and Textual Elements	12	9	21	26	17%
Total	**44**	**71**	**115**	**153**	

Note that four passages and their corresponding items are common to both the PIRLS and PIRLS Literacy assessments.

Developing the Materials for PIRLS, PIRLS Literacy, and ePIRLS 2016 Field Test Scoring Training

In order for field test scoring to occur immediately upon completion of data collection, it was necessary to prepare scoring training materials for the newly developed constructed response items in advance of the field test.

For PIRLS, to provide "grist" for these scoring materials, Australia, Canada (Ontario), England, Ireland, and Singapore administered the newly developed constructed response field test items in a small selection of classrooms with English-speaking students. Approximately 100 sample responses to each newly developed constructed response field test item were collected in October–November 2014.

For PIRLS Literacy, the participating countries either were not English-speaking countries or on a Southern Hemisphere school schedule. Thus, the TIMSS & PIRLS International Study Center worked with the Boston College department responsible for working with local school districts to administer the newly developed PIRLS Literacy constructed response items to a range of third grade classrooms in the Boston area. Approximately 50–100 responses to each item were collected in October–November 2014.

For ePIRLS, about 50 responses to each constructed response item were collected in November 2014 as part of the ePIRLS pilot to test the systems in advance of the field test. Approximately 50 sample responses for each item were collected from students in Australia, Ireland, and Canada (Ontario).

Exhibit 1.7 provides the number of constructed response items included in the effort to collect student responses for developing scoring training materials and the number of student responses collected.

Exhibit 1.7: Collecting Student Responses for Developing Field Test Scoring Training Materials

	PIRLS	PIRLS Literacy	ePIRLS
Passages/Tasks	12	10	6
Items			
Total	**114**	**88**	**71**
Responses per item (approx.)	100	30	50
Participants			
Countries	Australia, Canada (Ontario), England, Ireland, Singapore	TIMSS & PIRLS International Study Center	Australia, Ireland, Canada (Ontario)

Note that four passages and their corresponding items are common to both the PIRLS and PIRLS Literacy assessments.

A working group consisting of Marian Sainsbury and Liz Twist from NFER, Prue Anderson from ACER, Karen Wixson from the RDG, and Ina Mullis from the TIMSS & PIRLS International Study Center created sets of example and practice responses for 41 fourth grade PIRLS and PIRLS Literacy items. The example and practice response sets for each item included a scoring guide, approximately 8–10 example responses illustrating the categories in the scoring guide, and approximately 8–10 practice responses so that country representatives could practice making distinctions among categories and reach agreement about how to make consistent scoring decisions across countries. For ePIRLS, Marian Sainsbury and the TIMSS & PIRLS International Study Center used computer produced Excel sheets of responses to develop scoring guides and example responses.

The PIRLS 2016 NRCs and their scoring supervisors received scoring training for the field test constructed response items in February 2015 in Floriana, Malta as part of the 4th PIRLS 2016 NRC meeting. This training was conducted by the scoring training team, which included Julian Fraillon and Prue Anderson of ACER and Marian Sainsbury of NFER. At the scoring training sessions, the trainers explained the purpose of each item and read it aloud. The trainer then described the scoring guide, explaining each category and the rationale for the score given to each example response. After the country representatives scored the practice responses, the NRCs and the scoring training team discussed any inconsistencies in scoring. When necessary, the field test guides were clarified and sometimes categories were revised.

Finalizing the PIRLS, PIRLS Literacy, and ePIRLS 2016 Achievement Items

Subsequent to the field test, the TIMSS & PIRLS International Study Center analyzed the field test data and prepared almanacs containing summary item statistics for each field test item. The data almanac for an item contained, row by row for each country: the sample size, the item difficulty and discrimination, the percentage of students answering each option (multiple-choice) or in each score category (constructed response), the point-biserial correlation for each multiple-choice option or constructed response category, and the degree of scoring agreement for constructed response items.

The field test data were used by the TIMSS & PIRLS International Study Center, the RDG, and NRCs to assess the quality of the field test items. The TIMSS & PIRLS International Study Center staff members, together with external consultants, first reviewed the field test data to make an initial judgment about the quality of each item based on its measurement properties (item statistics). Items were eliminated from further consideration if they had poor measurement properties, such as being too difficult or easy or having low discrimination. Particular attention was paid to unusual item statistics in individual countries because these could indicate errors in translation.

After the item-by-item review, the TIMSS & PIRLS International Study Center staff collaborated with consultants to assemble a set of recommended passages with their item sets and ePIRLS tasks for review by the RDG. RDG members scrutinized the recommendations for the newly developed assessment materials, reviewing each passage and item set as well as scoring guides for content accuracy, clarity, and adherence to the frameworks. In addition, the newly developed passages and items were considered in relation to the trend passages and item sets for overall coherence as a complete assessment. The ePIRLS tasks and items were reviewed via computer. Five of the six ePIRLS tasks that were field tested were recommended for inclusion in the ePIRLS assessment.

NRCs had the opportunity to review the recommended materials in light of the field test results and within the security of their own countries. Each country also could check any unusual national results that might indicate translation errors and correct the translation as necessary or recommend revisions to accommodate translation. The 5^{th} NRC meeting held in Jyväskylä, Finland in August 2015 was devoted to reviewing all the recommended passages, tasks, and items for PIRLS, PIRLS Literacy, and ePIRLS. Following this meeting, the TIMSS & PIRLS International Study Center staff implemented revisions to the passages, tasks, and items as recommended by the NRCs. Final versions of the materials were distributed to the NRCs in August 2015.

Exhibit 1.8 includes descriptions of the PIRLS 2016 and PIRLS Literacy 2016 passages, including the newly developed passages for PIRLS 2016 and trend passages from PIRLS 2001, 2006, and 2011.

Exhibit 1.8: PIRLS 2016 Assessment Passages

Literary Passages	Informational Passages
PIRLS Passages	
Shiny Straw ■ – This animal story demonstrates heroism and the consequences of a reckless attitude.	**Leonardo Da Vinci** ◊ – This biographical text describes the inventions of Leonardo da Vinci and the ways that he was ahead of his time.
Macy and the Red Hen – This contemporary story portrays a complex character who meets a challenge when caring for a red hen.	**The Green Sea Turtle's Journey of a Lifetime** – This passage describes the life cycle of a female green sea turtle from the time she hatches from an egg to the time she lays her own eggs.
The Empty Pot * - This traditional tale set in China has a moral message about the importance of honesty.	**Where's the Honey?** * – This passage describes the relationship between the honeyguide bird and the Boran people in Africa using a combination of explanation, photographs, and graphic displays.
Oliver and the Griffin - In this fantasy story, a boy named Oliver meets an old griffin in a garden and decides to help him.	**Icelandic Horses** – This article describes the history and characteristics of Icelandic horses as they developed along with the people who lived near them.
Shared PIRLS/PIRLS Literacy Passages	
Flowers on the Roof ◊ – This contemporary story portrays friendship between the generations.	**Sharks** ■ – This article presents information about sharks in a variety of formats, using subheadings, a labeled diagram, and photographs.
Pemba Sherpa – This modern tale set in the Himalayan Mountains tells the story of a young girl determined to be a sherpa.	**How Did We Learn to Fly?** – This historical text explains how the modern airplane was developed.
PIRLS Literacy Passages	
Baghita's Perfect Orange * - This traditional tale set in Africa has a moral about greed and generosity.	**Training a Deaf Polar Bear** * – The passage describes how zookeepers worked with a polar bear that was found to be deaf.
The Pearl – This story about a young pearl merchant illustrates the power of home, friendship, and generosity above greed.	**African Rhinos & Oxpecker Birds** – This passage presents information about African rhinos and oxpecker birds and describes how the two animals depend on one another for food and survival.
The Summer My Father Was Ten * – In this thought-provoking story with a realistic contemporary setting, a boy is allowed to make amends for his thoughtless behavior.	**Ants** * – This article presents information about the lives of different types of ants, using subheadings, photographs, and diagrams.
Library Mouse – This story is about a mouse who lives in the library and inspires young children to be authors.	**Hungry Plant** – This scientific text describes the Venus Flytrap plant and explains how it captures insects for food.

◊ Passage from PIRLS 2001
■ Passage from PIRLS 2006
* Passage from PIRLS 2011

Exhibit 1.8: PIRLS 2016 Assessment Passages (Continued)

PIRLS 2016 Word Counts and Readability

Passage	Word Count	Flesch-Kincaid Grade Level
PIRLS Passages		
Shiny Straw	860	5.5
Macy and the Red Hen	913	4.4
The Empty Pot	767	4.9
Oliver and the Griffin	896	3.3
Leonardo Da Vinci	869	5.1
The Green Sea Turtle's Journey of a Lifetime	943	4.0
Where's the Honey?	870	3.2
Icelandic Horses	870	5.0
Shared PIRLS/PIRLS Literacy Passages		
Flowers on the Roof	811	2.8
Sharks	570	7.6
Pemba Sherpa	540	2.5
How Did We Learn to Fly?	514	6.3
PIRLS Literacy Passages		
Baghita's Perfect Orange	404	2.0
The Pearl	536	2.9
The Summer My Father Was Ten	484	4.0
Library Mouse	497	3.1
Training a Deaf Polar Bear	425	4.0
African Rhinos & Oxpecker Birds	449	4.7
Ants	415	2.9
Hungry Plant	509	3.5

The Flesch-Kincaid Grade Level Formula uses average syllables per word and average sentence length to produce a number that represents the US grade in which students can read the text.

Exhibit 1.9 includes descriptions of the ePIRLS tasks assessing online informational reading.

Exhibit 1.9: ePIRLS 2016 Assessment Tasks

Mars – In this science task, students learn what scientists know about Mars and investigate space exploration.

Dr. Elizabeth Blackwell – This biographical task is about the life and accomplishments of Elizabeth Blackwell, the first female doctor in both America and England.

Rainforests – This science task is about the plants and animals that live in the rainforest.

Zebra and Wildebeest Migration – Students learn about zebra and wildebeest migration through the Serengeti.

The Legend of Troy – This historical task is about the legend of Troy and archeological investigations of the ancient city.

Distribution of PIRLS 2016 Items by Reading Purpose and Comprehension Process

Exhibits 1.10 and 1.11 present the number of trend and newly developed items as well as the number of score points in the PIRLS 2016 assessments. The number of items represents the number of distinct questions in the assessment, while the number of score points represents the complexity and weight given to each item. Half the PIRLS and PIRLS Literacy items are based on literary passages and half are based on informational passages. ePIRLS assesses reading for information, but in an online environment.

Exhibit 1.10: PIRLS 2016 Achievement Items by Reading Purpose

Reading Purpose	Number of Passages/ Tasks	Number of Trend Items in PIRLS 2016	Percentage of Trend Score Points	Number of New Items in PIRLS 2016	Percentage of New Score Points	Total Items	Achieved Percentage of Score Points	Target Percentage of Score Points
				PIRLS				
Literary	6	44 (55)	49%	46 (58)	51%	90 (113)	51%	50%
Informational	6	37 (51)	46%	48 (59)	54%	85 (110)	49%	50%
Total	**12**	**81 (106)**		**94 (117)**		**175 (223)**		
				PIRLS Literacy				
Literary	6	43 (48)	45%	50 (59)	55%	93 (107)	50%	50%
Informational	6	40 (51)	49%	50 (54)	51%	90 (105)	50%	50%
Total	**12**	**83 (99)**		**100 (113)**		**183 (212)**		
				ePIRLS				
Informational	5	0 (0)	0%	91 (112)	100%	91 (112)	100%	100%
Total	**5**	**0 (0)**		**91 (112)**		**91 (112)**		

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.
Note that four passages and their corresponding items are common to both the PIRLS and PIRLS Literacy assessments.

Exhibit 1.11: PIRLS 2016 Achievement Items by Comprehension Process

Comprehension Process	Number of Trend Items in PIRLS 2016	Percentage of Trend Score Points	Number of New Items in PIRLS 2016	Percentage of New Score Points	Total Items	Achieved Percentage of Score Points	Target Percentage of Score Points
PIRLS							
Focus on & Retrieve Explicitly Stated Information	18 (21)	37%	32 (36)	63%	50 (57)	26	20
Make Straightforward Inferences	28 (30)	51%	25 (29)	49%	53 (59)	26	30
Interpret & Integrate Ideas and Information	24 (42)	53%	23 (37)	47%	47 (79)	35	30
Evaluate & Critique Content and Textual Elements	11 (13)	46%	14 (15)	54%	25 (28)	13	20
Total	**81 (106)**		**94 (117)**		**175 (223)**		
PIRLS Literacy							
Focus on & Retrieve Explicitly Stated Information	31 (36)	40%	51 (55)	60%	82 (91)	43	50
Make Straightforward Inferences	27 (27)	50%	25 (27)	50%	52 (54)	25	25
Interpret & Integrate Ideas and Information	15 (26)	53%	16 (23)	47%	31 (49)	23	25
Evaluate & Critique Content and Textual Elements	10 (10)	56%	8 (8)	44%	18 (18)	8	
Total	**83 (99)**		**100 (113)**		**183 (212)**		
ePIRLS							
Focus on & Retrieve Explicitly Stated Information	0 (0)	0%	22 (23)	100%	22 (23)	21	20
Make Straightforward Inferences	0 (0)	0%	27 (31)	100%	27 (31)	28	30
Interpret & Integrate Ideas and Information	0 (0)	0%	23 (38)	100%	23 (38)	34	30
Evaluate & Critique Content and Textual Elements	0 (0)	0%	19 (20)	100%	19 (20)	18	20
Total	**0 (0)**		**91 (112)**		**91 (112)**		

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.
Note that four passages and their corresponding items are common to both the PIRLS and PIRLS Literacy assessments.

Distribution of PIRLS Item Formats within Reading Purposes and Comprehension Processes

As described in the *PIRLS 2016 Assessment Framework*, up to half of the total number of score points represented by all the questions come from multiple-choice items. Most PIRLS multiple-choice items are worth one score point, although some compound multiple-choice items are worth two score points. The 2-point compound multiple-choice items are scored as all parts answered correctly as fully correct (2 score points), and most parts answered correctly as partially correct (1 score point). Constructed response items generally are worth one, two, or three score points depending on the degree of complexity involved. The 1-point constructed response items are scored as correct (1 score point) or incorrect (0 score points), whereas 2-point constructed response items are scored as fully correct (2 score points), partially correct (1 score point), or incorrect (0 score points), and 3-point constructed response items are scored as fully correct (3 score points), partially correct (1 or 2 score points), or incorrect (0 score points). Fully correct responses show a complete or deeper understanding of a task while partially correct responses demonstrate only a partial understanding of the concepts embodied in the task.

Exhibits 1.12 and 1.13 display the number of passages or tasks and items (and score points) by item format for each purpose and comprehension process.

Exhibit 1.12: PIRLS 2016 Achievement Items by Reading Purpose and Item Format

Reading Purpose	Number of Passages/ Tasks	Multiple-Choice Items		Constructed Response Items			Total Items	Percentage of Score Points
		Four Response Options	Compound	1 Point	2 Points	3 Points		
PIRLS								
Literary	6	46 (46)	0 (0)	25 (25)	15 (30)	4 (12)	90 (113)	51%
Informational	6	40 (40)	0 (0)	24 (24)	17 (34)	4 (12)	85 (110)	49%
Total	**12**	**86 (86)**	**0 (0)**	**49 (49)**	**32 (64)**	**8 (24)**	**175 (223)**	
Achieved Percentage of Score Points		**39%**		**61%**				
Target Percentage of Score Points		**40%**		**60%**				
PIRLS Literacy								
Literary	6	47 (47)	0 (0)	33 (33)	12 (24)	1 (3)	93 (107)	50%
Informational	6	43 (43)	1 (2)	34 (34)	10 (20)	2 (6)	90 (105)	50%
Total	**12**	**90 (90)**	**1 (2)**	**67 (67)**	**22 (44)**	**3 (9)**	**183 (212)**	
Achieved Percentage of Score Points		**43%**		**57%**				
Target Percentage of Score Points		**40%**		**60%**				
ePIRLS								
Literary	0	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0%
Informational	5	36 (36)	4 (8)	37 (37)	11 (22)	3 (9)	91 (112)	100%
Total	**5**	**36 (36)**	**4 (8)**	**37 (37)**	**11 (22)**	**3 (9)**	**91 (112)**	
Achieved Percentage of Score Points		**39%**		**61%**				
Target Percentage of Score Points		**40%**		**60%**				

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.
Note that four passages and their corresponding items are common to both the PIRLS and PIRLS Literacy assessments.

Exhibit 1.13: PIRLS 2016 Achievement Items by Comprehension Process and Item Format

Comprehension Process	Multiple-Choice Items		Constructed Response Items			Total Items	Percentage of Score Points
	Four Response Options	Compound	1 Point	2 Points	3 Points		
PIRLS							
Focus on and Retrieve Explicitly Stated Information	25 (25)	0 (0)	18 (18)	7 (14)	0 (0)	50 (57)	26%
Make Straightforward Inferences	35 (35)	0 (0)	12 (12)	6 (12)	0 (0)	53 (59)	26%
Interpret and Integrate Ideas and Information	11 (11)	0 (0)	12 (12)	16 (32)	8 (24)	47 (79)	35%
Evaluate and Critique Content and Textual Elements	15 (15)	0 (0)	7 (7)	3 (6)	0 (0)	25 (28)	13%
Total	**86 (86)**	**0 (0)**	**49 (49)**	**32 (64)**	**8 (24)**	**175 (223)**	
Achieved Percentage of Score Points	**39%**		**61%**				
Target Percentage of Score Points	**40%**		**60%**				
PIRLS Literacy							
Focus on and Retrieve Explicitly Stated Information	30 (30)	0 (0)	43 (43)	9 (18)	0 (0)	82 (91)	43%
Make Straightforward Inferences	35 (35)	0 (0)	15 (15)	2 (4)	0 (0)	52 (54)	25%
Interpret and Integrate Ideas and Information	8 (8)	1 (2)	8 (8)	11 (22)	3 (9)	31 (49)	23%
Evaluate and Critique Content and Textual Elements	17 (17)	0 (0)	1 (1)	0 (0)	0 (0)	18 (18)	8%
Total	**90 (90)**	**1 (2)**	**67 (67)**	**22 (44)**	**3 (9)**	**183 (212)**	
Achieved Percentage of Score Points	**43%**		**57%**				
Target Percentage of Score Points	**40%**		**60%**				
ePIRLS							
Focus on and Retrieve Explicitly Stated Information	10 (10)	0 (0)	11 (11)	1 (2)	0 (0)	22 (23)	21%
Make Straightforward Inferences	12 (12)	0 (0)	11 (11)	4 (8)	0 (0)	27 (31)	28%
Interpret and Integrate Ideas and Information	3 (3)	4 (8)	8 (8)	5 (10)	3 (9)	23 (38)	34%
Evaluate and Critique Content and Textual Elements	11 (11)	0 (0)	7 (7)	1 (2)	0 (0)	19 (20)	18%
Total	**36 (36)**	**4 (8)**	**37 (37)**	**11 (22)**	**3 (9)**	**91 (112)**	
Achieved Percentage of Score Points	**39%**		**61%**				
Target Percentage of Score Points	**40%**		**60%**				

Score points are shown in parentheses.
Because percentages are rounded to the nearest whole number, some totals may appear inconsistent.
Note that four passages and their corresponding items are common to both the PIRLS and PIRLS Literacy assessments.

PIRLS 2016 Constructed Response Scoring Training

In preparation for the main data collection scoring training, some PIRLS 2016 scoring guides were further refined or clarified based on the results of the field test. This included a thorough review of the field test scoring training materials to ensure that the student responses were still suitable for the updated scoring guides. In some cases, example and practice sets used in the field test were expanded to further illustrate particular aspects of a scoring guide. For PIRLS/PIRLS Literacy 2016 scoring training, the example and practice paper training sets included those used in PIRLS 2011 for the trend items and the updated training sets for the newly developed items selected for PIRLS 2016, resulting in 42 example and practice paper sets for PIRLS and 24 for PIRLS Literacy. Scoring training materials were developed for 8 ePIRLS items.

To provide scoring training for all the countries participating in PIRLS 2016, the TIMSS & PIRLS International Study Center conducted two training sessions. First, the NRCs for Southern Hemisphere countries and their scoring supervisors received PIRLS and PIRLS Literacy scoring training in November 2015 in Buenos Aires, Argentina. (No Southern Hemisphere countries participated in ePIRLS.) NRCs for Northern Hemisphere countries and their scoring supervisors received scoring training in March 2016 in Hong Kong SAR as part of the 6th PIRLS 2016 NRC meeting.

Exhibit 1.14 shows the number of participants in the two scoring training sessions.

Exhibit 1.14: PIRLS 2016 Scoring Training Participation

Participants	Southern Hemisphere	Northern Hemisphere
Number of Countries	6	49
Number of Benchmarking Entities	2	10
Number of Country Representatives	29	119

The Process Following Instrument Development

In general, after the participating countries received the international version of the assessment instruments, they began the process of translation and cultural adaptation (some adaptation to local usage typically is necessary even in English-speaking countries) and production of the materials for printing. At the same time, countries made final arrangements for data collection, including the host of activities necessary to obtain school participation, implement test administration, and score the responses to the tests and questionnaires (see following chapters).

Reference

Kincaid, J.P., Fishburne, R.P., Jr., Rogers, R.L., & Chissom, B.S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel.* (Research Branch Report 8-75). Millington, TN: Naval Technical Training Command Millington Tennessee Research Branch.

CHAPTER 2

Developing the PIRLS 2016 Context Questionnaires

Martin Hooper
Bethany Fishbein

To provide insight into students' contexts for learning across participating countries, PIRLS context questionnaires are completed by students and their parents, teachers, and principals. National Research Coordinators (NRCs) from participating countries document national policies by completing a curriculum questionnaire.

The context questionnaire results form the basis for seven of the ten chapters of the *PIRLS 2016 International Results in Reading* report and one of the four chapters of the *ePIRLS International Results in Online Informational Reading* report, and the descriptive data collected through the PIRLS Curriculum Questionnaire complement each country's chapter in the *PIRLS 2016 Encyclopedia*. The data are also made available through the PIRLS 2016 International Database, providing data that researchers can use for secondary analysis.

This chapter documents the PIRLS 2016 questionnaire development process. Information on the analysis of the context questionnaire scales can be found in Chapter 14.

Development Process for the PIRLS 2016 Context Questionnaires

Developing the PIRLS 2016 context questionnaires was a collaborative process involving multiple rounds of reviews by staff at the TIMSS & PIRLS International Study Center, experts on the PIRLS 2016 Questionnaire Development Group (QDG), and the NRCs from the participating countries. In broad strokes, the PIRLS 2016 context questionnaire development process for the student, home, school, and teacher questionnaires included:

- Updating the context questionnaire framework for 2016
- Developing new context questionnaire items and modifying existing items by staff at the TIMSS & PIRLS International Study Center
- Reviewing and revising successive draft questionnaires by the QDG and NRCs
- Administering the PIRLS 2016 field test

- Using the field test results to refine the questionnaires by staff at the TIMSS & PIRLS International Study Center and the QDG
- Final review by NRCs

Developing the Curriculum Questionnaire followed a collaborative process similar to other PIRLS questionnaires, including identifying important framework topics, developing questionnaire items, and undergoing reviews by the QDG and NRCs.

Exhibit 2.1 presents the PIRLS 2016 context questionnaire development schedule. The development process was directed and managed by the staff of the TIMSS & PIRLS International Study Center at Boston College, including Executive Directors Ina V.S. Mullis and Michael O. Martin, and the PIRLS Questionnaire Coordinator, Martin Hooper. NRCs had an essential role in updating the questionnaires, providing feedback and ideas at NRC meetings. The QDG made major contributions in updating the PIRLS 2016 questionnaires with the 1st QDG meeting focused on developing PIRLS items/scales, and the 2nd meeting focused on refining the questionnaires in light of the field test results. Exhibit 2.2 lists the members of the QDG.

PIRLS 2016 included PIRLS Literacy, a less difficult version of PIRLS, and ePIRLS—a computer-based assessment of online informational reading. Countries participating in PIRLS Literacy administered the PIRLS questionnaires. All students taking ePIRLS also took PIRLS, and these students were administered a short ePIRLS questionnaire in addition to the PIRLS questionnaire.

Exhibit 2.1: PIRLS 2016 Context Questionnaire Development Schedule

Date(s)		Group and Activity
February	2013	NRCs reviewed PIRLS 2011 context questionnaires and provided ideas for new questionnaire topics at the 1st NRC meeting (Hamburg, Germany)
June	2013	1st meeting of the Questionnaire Development Group (QDG) to develop the PIRLS 2016 questionnaires (Singapore). Meeting was held jointly with the TIMSS Questionnaire Item Review Committee (QIRC)
July–August	2013	TIMSS & PIRLS International Study Center revised the draft context questionnaires to incorporate QDG/QIRC feedback and drafted the PIRLS 2016 Context Questionnaire Framework chapter
September	2013	NRCs reviewed draft PIRLS 2016 context questionnaires and the draft PIRLS 2016 Context Questionnaire Framework chapter at the 2nd NRC meeting (Portorož, Slovenia)
October	2013	TIMSS & PIRLS International Study Center finalized the PIRLS 2016 Context Questionnaire Framework chapter incorporating NRC feedback
October–April	2013-2014	TIMSS & PIRLS International Study Center updated the draft PIRLS 2016 context questionnaires incorporating NRC feedback
November	2013	TIMSS & PIRLS International Study Center published *PIRLS 2016 Assessment Frameworks, 1st Edition*, which includes the chapter on the Context Questionnaire Framework
May	2014	NRCs reviewed and approved the proposed field test context questionnaires for PIRLS at the 3rd NRC meeting (Dublin, Ireland)
June–July	2014	TIMSS & PIRLS International Study Center finalized field test context questionnaire instruments
July	2014	TIMSS & PIRLS International Study Center provided field test context questionnaires to NRCs
February	2015	*PIRLS 2015 Assessment Framework, 2nd Edition* published online
March–April	2015	Countries conducted PIRLS 2016 field test
April–May	2015	Countries submitted field test data for analysis and review
June	2015	TIMSS & PIRLS International Study Center conducted a review of field test results
July	2015	QDG reviewed questionnaire field test data and the draft PIRLS 2016 Curriculum Questionnaire at 2nd QDG meeting (Hamburg, Germany)
August	2015	NRCs reviewed and approved context questionnaires for PIRLS and ePIRLS 2016 data collection as well as the PIRLS 2016 Curriculum Questionnaire at 5th NRC meeting (Jyväskylä, Finland)
August	2015	TIMSS & PIRLS International Study Center distributed PIRLS 2016 and ePIRLS 2016 context questionnaire instruments for data collection to NRCs for translation
October–December	2015	Southern Hemisphere countries conducted PIRLS 2016 data collection
March–June	2016	Northern Hemisphere countries conducted PIRLS 2016 data collection
March-August	2016	PIRLS 2016 Curriculum Questionnaire administered online to NRCs

Exhibit 2.2: PIRLS 2016 Questionnaire Development Group (QDG)

Joanne Latourelle
Sanction des Études
Ministère de l'Education, et de L'EnseignementSupérieur
Canada

Hwa Wei Ko
Graduate Institute of Learning and Instruction
National Central University
Chinese Taipei

Marc Colmant
Direction de l'Évaluation, de la Prospective et de la Performance (DEPP)
Ministère de l'Éducation Nationale
France

Maryam A. Al-Ostad
National Centre for Education Development
Kuwait

Megan Chamberlain
Comparative Education Research Unit
Ministry of Education
New Zealand

João Maroco
Instituto de Avaliação Educativa, I. P.
Portugal

Sarah Howie
Centre for Evaluation and Assessment (CEA)
University of Pretoria
South Africa

Background of PIRLS 2016 Context Questionnaire Development

Similar to the development process for the PIRLS 2016 achievement booklets (see Chapter 1), questionnaire development balanced the dual purposes of maintaining continuity with previous assessments and evolving to reflect the current contexts for student learning. Following from this, the PIRLS 2016 questionnaires were based on the questionnaires from PIRLS 2011 and informed by developments for TIMSS 2015, with updates as appropriate to align the questionnaires with more recent research on favorable contexts for learning to read.

In 2011, the TIMSS and PIRLS cycles coincided, and 34 countries chose to administer both TIMSS and PIRLS to the same fourth grade students. Accordingly, the TIMSS 2011 and PIRLS 2011 questionnaires were developed in tandem (see *Methods and Procedures in TIMSS and PIRLS 2011* for details). Overall, this joint development process produced a synergy that led to advancements in questionnaire development for both projects, and shared items across TIMSS and PIRLS 2011 allowed results to be compared across projects.

PIRLS 2016 made an effort to maintain the consistency with TIMSS by holding the 1st meeting of the QDG with its TIMSS equivalent—the Questionnaire Item Review Committee (QIRC). Because TIMSS is on a four-year cycle and PIRLS is on a five-year cycle, much of the

TIMSS 2015 development occurred in advance of the PIRLS 2016 development, allowing PIRLS 2016 to capitalize on improvements made to the TIMSS 2015 questionnaires. As such, the PIRLS 2016 Context Questionnaire Framework built upon the research conducted for the TIMSS 2015 framework, and the PIRLS 2016 questionnaire development benefitted from revisions to overlapping TIMSS/PIRLS questionnaire items made at TIMSS NRC meetings. PIRLS 2016 development also was informed by results from the TIMSS 2015 field test.

A major methodological innovation in PIRLS 2011 (and TIMSS 2011) was using context questionnaire scales to measure key educational research topics (Martin, Mullis, Foy, & Arora, 2012). To improve scales for PIRLS 2016, questionnaire development focused on writing items to strengthen the measurement properties of the PIRLS 2011 scales as well as developing new scales to measure emerging areas of educational research.

Updating the PIRLS 2016 Context Questionnaire Framework

The PIRLS 2016 Context Questionnaire Framework, Chapter 2 of the *PIRLS 2016 Assessment Framework*, provided the foundation for updating the PIRLS context questionnaires for 2016. The chapter presents a review of the educational research that identifies key context questionnaire topics and gives the rationale for asking about these topics within the 2016 questionnaires.

At the 1st NRC meeting in February 2013 in Hamburg, Germany, NRCs described topics they thought should be covered in the PIRLS 2016 questionnaires, including which PIRLS 2011 topics should be retained to measure trends. Taking into account feedback garnered in the meeting and insights from the drafting of the TIMSS 2015 framework, the PIRLS Questionnaire Coordinator conducted a literature review and drafted the PIRLS 2016 Context Questionnaire Framework. Because the primary purpose of the context questionnaires is to identify factors that may contribute to differences in achievement within and between countries, the framework focuses on topics in educational research found to be related to achievement across a variety of settings and contexts.

The NRCs reviewed the draft framework chapter at the 2nd NRC meeting in September 2013 in Portorož, Slovenia. Staff at the TIMSS & PIRLS International Study Center refined the draft based upon the recommendations received at the meeting and published the *PIRLS 2016 Assessment Framework* online in November 2013, with printed copies distributed thereafter. A second edition of the framework was published in February 2015, which included updates to the PIRLS Literacy assessment design.

Field Test Questionnaire Development

With the draft Context Questionnaire Framework at hand, staff at the TIMSS & PIRLS International Study Center focused the questionnaire development process on improving and expanding the PIRLS context questionnaire scales and updating items to align with more recent technological innovations.

For many of the scales retained from PIRLS 2011, modifications for 2016 focused on increasing the number of items to optimize reliability and content coverage. For example, a number of new items were written for the *School Emphasis on Academic Success* scale, with item development influenced by existing scales in the academic optimism literature (Hoy, Hoy, & Kurz, 2008; McGuigan & Hoy, 2006; Wu, Hoy, & Tarter, 2013). New items asking teachers about their strategies for engaging students were revamped, with item development influenced by Applebee, Langer, Nystrand, and Gamoran (2003). Additional items were also included for the student engagement scales, with one item sourced from Fauth, Decristan, Rieser, Klieme, and Büttner (2014).

Staff at the TIMSS & PIRLS International Study Center worked with the PIRLS QDG/ TIMSS QIRC at their joint meeting in June 2013 to recast a number of scales. For instance, the QDG and QIRC revamped the *Teacher Job Satisfaction* scale to integrate insights gained from the Utrecht Work Engagement Scale (Schaufeli, Bakker, & Salanova, 2006). The questionnaire committees also suggested a new item for the *Parents Like Reading* scale, sourced from PISA 2000 (OECD, 2000).

Updating questionnaires to "keep up with the times" was an essential part of the 2016 development process. Staff at the TIMSS & PIRLS International Study Center worked with the QIRC and QDG to ensure that the questionnaires included items on the availability of prevalent digital resources for education such as ebooks, tablets, and interactive whiteboards.

Finally, staff at the TIMSS & PIRLS International Study Center developed a short ePIRLS student questionnaire to focus on students' experiences using computers and finding and reading information on the Internet as well as their self-efficacy using computers, typing, and finding information on the Internet.

Prior to the field test, the PIRLS NRCs reviewed draft PIRLS 2016 questionnaires at their 2nd NRC meeting in September 2013 in Portorož, Slovenia, as well as at their 3rd NRC meeting in May 2014 in Dublin, Ireland. The ePIRLS questionnaire was also reviewed at the 3rd NRC meeting.

Review Field Test Results and Refine Questionnaires for Data Collection

PIRLS 2016 countries administered an ambitious field test, eliciting questionnaire data from 64,873 students, 62,716 parents, 1,840 school principals, and 3,287 teachers from the 49 countries and seven benchmarking entities for PIRLS and across seven countries and one benchmarking entity for PIRLS Literacy. The ePIRLS field test questionnaire was administered to 13,701 students from 15 countries as well as five benchmarking entities.

Following field test administration, staff at the TIMSS & PIRLS International Study Center produced data almanacs and scale summaries to facilitate the review of the field test data:

- Data almanacs document for each country the use of response categories for each context questionnaire item as well each item's relationship with achievement
- Scale summaries detail each scale's reliability, dimensionality, fit to the item response theory model, and relationship with achievement in each country

In June 2015, staff at the TIMSS & PIRLS International Study Center reviewed the field test context questionnaire results, proposing revisions to the QDG. At their 2nd meeting in July 2014, the QDG accepted many of the recommendations and suggested a few additional changes. In August 2015 at their 5th meeting, NRCs reviewed the final draft questionnaires and accepted the questionnaires with a few minor revisions. Following the NRC meeting, staff at the TIMSS & PIRLS International Study Center implemented the revisions and posted the final PIRLS instruments on August 27, 2015, so that countries could begin the translation process.

Developing the PIRLS 2016 Curriculum Questionnaire

The PIRLS Curriculum Questionnaire complements the student, teacher, school, and home questionnaires by collecting information from NRCs about country-level contexts. The Curriculum Questionnaire covers each country's reading curriculum, goals and standards for instruction, and other national or regional policies such as the preprimary education process and the teacher education process.

Similar to the other PIRLS 2016 questionnaires, the process for updating the PIRLS Curriculum Questionnaire started with the PIRLS 2016 Context Questionnaire Framework. Then, the QDG identified the information from the PIRLS 2011 Curriculum Questionnaire and the TIMSS 2015 Curriculum Questionnaires that they thought was useful to continue collecting. Based on the framework and QDG feedback, staff at the TIMSS & PIRLS International Study Center updated the PIRLS 2016 Curriculum Questionnaire for review by NRCs at their 5th meeting in August 2015. Following the NRC meeting, staff at the TIMSS & PIRLS International Study Center finalized the questionnaire, incorporating the suggestions that emerged from the meeting. NRCs completed the online Curriculum Questionnaire between March 30, 2016 and August 31, 2016.

References

Applebee, A.N., Langer, J.A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal*, *40*(3), 685-730.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1–9.

Hoy, A.W., Hoy, W.K., & Kurz, N.M. (2008). Teacher's academic optimism: The development and test of a new construct. *Teaching and Teacher Education*, *24*, 821–835.

Martin, M.O., Mullis, I.V.S., Foy, P., & Arora, A. (2012). Creating and interpreting the TIMSS and PIRLS 2011 context questionnaire scales. In M.O. Martin & I.V.S. Mullis (Eds.), *Methods and Procedures in TIMSS and PIRLS 2011*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/methods/pdf/TP11_Context_Q_Scales.pdf

McGuigan, L. & Hoy, W.K. (2006). Principal leadership: Creating a culture of academic optimism to improve achievement for all students. *Leadership and Policy in Schools*, 5, 203–229.

Organisation for Economic Co-Operation and Development (OECD) (2000). Student Questionnaire. Programme for International Student Assessment. Retrieved from OECD website: http://www.oecd.org/pisa/pisaproducts/database-pisa2000.htm

Schaufeli, W.B., Bakker, A.B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, *66*(4), 701–716.

Wu, J.H., Hoy, W.K., & Tarter, C.J. (2013). Enabling school structure, collective responsibility, and a culture of academic optimism: Toward a robust model of school performance in Taiwan. *Journal of Educational Administration*, *51*(2), 176–193.

PIRLS
2016

Methods and Procedures in PIRLS 2016

Sampling

CHAPTER 3

Sample Design in PIRLS 2016

Sylvie LaRoche
Marc Joncas
Pierre Foy

Introduction

PIRLS is designed to provide valid and reliable measurement of trends in student achievement in countries around the world, while keeping to a minimum the burden on schools, teachers, and students. The PIRLS program employs rigorous school and classroom sampling techniques so that achievement in the student population as a whole may be estimated accurately by assessing just a sample of students from a sample of schools. PIRLS assesses reading achievement at fourth grade. 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.

PIRLS employs a two-stage random sample design, with a sample of schools drawn as a first stage and one or more intact classes of students selected from each of the sampled schools as a second stage. Intact classes of students are sampled rather than individuals from across the grade level or of a certain age because PIRLS pays particular attention to students' curricular and instructional experiences, and these typically are organized on a classroom basis. Sampling intact classes also has the operational advantage of less disruption to the school's day-to-day business than individual student sampling.

National Sampling Plan

Each country participating in PIRLS needs a plan for defining its national target population and applying the PIRLS sampling methods to achieve a nationally representative sample of schools and students. The development and implementation of the national sampling plan is a collaborative exercise involving the country's National Research Coordinator (NRC) and PIRLS sampling experts.

Statistics Canada is responsible for advising the National Research Coordinator on all sampling matters and for ensuring that the national sampling plan conforms to the PIRLS standards. In cooperation with sampling staff from IEA Hamburg, Statistics Canada works with the

National Research Coordinator to select the national school sample(s) and produce all supporting documentation for tracking the sampled schools. This includes ensuring that the school sampling frame (the school population list from which the school sample is drawn) provided by the National Research Coordinator is complete and satisfactory; checking that categories of excluded students are clearly defined, justified, and kept to a minimum; assisting the National Research Coordinator in determining the sample size and a stratification plan that will meet both international and national objectives; and drawing a national sample of schools. When sampling has been completed and all data collected, Statistics Canada documents population coverage and school and student participation rates and constructs appropriate sampling weights for use in analyzing and reporting the results.

The TIMSS & PIRLS International Study Center, in cooperation with Statistics Canada and IEA Hamburg, provides National Research Coordinators with a series of manuals to guide them through the sampling process. More specifically, *PIRLS 2016 Survey Operations Procedures Unit 1: Sampling Schools and Obtaining their Cooperation* describes the steps involved in defining the national target population and selecting the school sample, and *PIRLS 2016 Survey Operations Procedures Unit 3: Contacting Schools and Sampling Classes for Data Collection* describes the procedure for sampling classes within the sampled schools and making preparations for conducting the assessments. Within-school sampling procedures for the field test are documented in *PIRLS 2016 Survey Operations Procedures Unit 2: Preparing for and Conducting the Field Test*. More information on the Survey Operations Units can be found in Chapter 6 of this publication.

The PIRLS National Research Coordinator is responsible for providing Statistics Canada with all information and documentation necessary to conduct the national sampling, and for conducting all sampling operations in the country. In particular, the NRC is expected to identify the grade that corresponds to the international target population; create a sampling frame by listing all schools in the population that have classes with students in the target grade; determine national population coverage and exclusions, in accordance with the PIRLS international guidelines; work with Statistics Canada to develop a national sampling plan and identify suitable stratification variables, ensuring that these variables are present and correct for all schools; contact all sampled schools and secure their participation; keep track of school participation and the use of replacement schools; and conduct all within-school sampling of classes. Each NRC is required to complete a series of sampling forms documenting the completion of each of these tasks.

A crucial feature of each international meeting of National Research Coordinators is a one-to-one meeting between each NRC and sampling staff at Statistics Canada and IEA Hamburg. At these meetings, each step of the sampling process is documented and reviewed in detail, and NRCs have the opportunity to raise issues and ask questions about their national situation and any challenges they face. Statistics Canada consults with the TIMSS & PIRLS International Study

Center and the International Sampling Referee, as necessary, to resolve issues and questions. Final approval of PIRLS national sampling plans is the responsibility of the TIMSS & PIRLS International Study Center, based upon the advice of Statistics Canada and the International Sampling Referee.

Defining the Target Population

As an international study of the comparative effects of education on student achievement in reading literacy, PIRLS defines its international target population in terms of the amount of schooling students have received. The number of years of formal schooling is the basis of comparison among participating countries. Thus, the PIRLS international target population is all students in their fourth year of formal schooling. UNESCO's *International Standard Classification of Education ISCED 2011* (UNESCO, 2012) provides an internationally accepted classification scheme for describing levels of schooling across countries. The ISCED system describes the full range of schooling, from pre-primary (Level 0) to the doctoral level (Level 8). ISCED Level 1 corresponds to primary education or the first stage of basic education. The first year of Level 1 "coincides with the transition point in an education system where systematic teaching and learning in reading, writing and mathematics begins" (UNESCO, 2012, p. 30). Four years after this would be the target grade for PIRLS, and is the fourth grade in most countries. However, given the cognitive demands of the assessments, PIRLS wants to avoid assessing very young students. Thus, PIRLS recommends assessing the next higher grade (i.e., fifth grade) if the average age at the time of testing would be less than 9.5 years.

The PIRLS target population of students is defined as follows:

> All students enrolled in the grade that represents four years of schooling counting from the first year of ISCED Level 1, providing the mean age at the time of testing is at least 9.5 years.

All students enrolled in the target grade, regardless of their age, belong to the international target population and should be eligible to participate in PIRLS. Because students are sampled in two stages, first by randomly selecting a school and then randomly selecting a class from within the school, it is necessary to identify all schools in which eligible students are enrolled. Essentially, eligible schools for PIRLS are those that have any students enrolled in the target grade, regardless of type of school. All schools of all educational sub-systems that have students learning full-time in the target grade are part of the international target population, including schools that are not under the authority of the national Ministry of Education.

National Target Population

For most countries, the target grade for PIRLS is the fourth grade. However, because educational systems vary in structure and in policies and practices with regard to age of starting school and promotion and retention, there are differences across countries in how the target grade is labelled and in the average age of students. To ensure that the appropriate national target grade is selected, each NRC completes Sampling Form 1, which identifies the target grade, the country's name for the grade, and the average age of students in that grade at the time of data collection. An example of a completed Sampling Form 1 is presented in Exhibit 3.1.

Exhibit 3.1: Example of Sampling Form 1

Sampling Form 1 General Information

See Section 2 of the Survey Operation Manual Unit 1

PIRLS 2016 Participant: < Name of the Country >

National Research Coordinator : < Name of the NRC >

1. Please indicate studies in which your country plans to participate along with the targeted grade(s), name(s), and expected average age of students at the time of testing:

	Target Grade	Name of the Target Grade	Average Age
PIRLS	4	Primary 4	9.7
PIRLS Literacy			

ePIRLS	Yes

2. Specify the usual start and end date of the school year.

Start of school year : *(YYYY-MM-DD)* September 1, 2015

End of school year: *(YYYY-MM-DD)* June 21, 2016

3. Specify the expected testing periods of surveying for the Field Test and the Data Collection.

Expected testing period (Field Test): 15 April 2015

Expected testing period (Data Collection): 13-14 April 2016

4. Will you request that Statistics Canada and the IEA DPC select your school sample(s)? *(Click in box and on right arrow to see drop down menu)*

Yes

5. Specify the language(s) in which the survey will be administered. If your response differs for the Field Test and the Data Collection, please split your response by survey phase.

English

6. Describe the grade structure through ISCED level 1 (primary education or the first stage of basic education) and level 2 (basic or lower secondary education) in your country.

Grades 1 to 6 , Primary schools
primary and lower secondary education are generally found in the same schools. Some schools also offer only primary education

7. Describe the age and birth date rules for entering ISCED level 1 in your country.

Children must enter school (grade 1) in the autumn of the year in which they have their sixth birthday

National Coverage and Exclusions

PIRLS is designed to describe and summarize student achievement across the entire target grade, and so it is very important that the national target population aims for comprehensive coverage of eligible students. However, in some cases, political, organizational, or operational factors make complete national coverage difficult to attain. Thus, in some rare situations, certain groups of schools and students may have to be excluded from the national target population. For example, it may be that a particular geographical region, educational sub-system, or language group cannot be covered. Such exclusion of schools and students from the target population is referred to as reduced population coverage.

Even countries with complete population coverage find it necessary to exclude at least some students from the target population because they attend very small schools, have intellectual or functional disabilities, or are non-native language speakers. Such students may be excluded at the school level (i.e., the whole school is excluded) or within the school on an individual basis.

School Level Exclusions. Although it is expected that very few schools will be excluded from the national target population, NRCs are permitted to exclude schools on the following grounds when they consider it necessary:

- Inaccessibility due to their geographically remote location
- Extremely small size (e.g., four or fewer students in the target grade)
- Offering a grade structure, or curriculum, radically different from the mainstream educational system
- Providing instruction solely to students in the student-level exclusion categories listed below (e.g., catering only to special needs students)

Student Level Exclusions. The international within-school exclusion rules are specified as follows:

- Students with functional disabilities — These are students who have physical disabilities such that they cannot perform in the PIRLS testing situation. Students with functional disabilities who are able to perform should be included in the testing.
- Students with intellectual disabilities — These are students who are considered, in the professional opinion of the school principal, or by other qualified staff members, to have intellectual disabilities or who have been tested as such. This includes students who are emotionally or mentally unable to follow even the general instructions of the test. Students should not be excluded solely because of poor academic performance or normal disciplinary problems. It should be noted that students with dyslexia, or other such learning disabilities, should be accommodated in the test situation if possible, rather than excluded.

- Non-native language speakers — These are students who are unable to read or speak the language(s) of the test and would be unable to overcome the language barrier in the test situation. Typically, a student who has received less than one year of instruction in the language(s) of the test should be excluded.

Because disability criteria vary from country to country, NRCs are asked to translate the PIRLS international exclusion standards into the local equivalent. Students should be considered for exclusion strictly in accordance with the international standards. If a sampled school contains a class consisting entirely of students from one of the exclusion categories, such a class is excluded prior to classroom sampling.

NRCs understand that exclusion rates must be kept to a minimum in order that national samples accurately represent the national target population.

- The overall number of excluded students must not account for more than 5% of the national target population of students in a country. The overall number includes both school-level and within-school exclusions.
- The number of students excluded because they attend very small schools must not account for more than 2% of the national target population of students.

To document population coverage and exclusions, each NRC completes Sampling Form 2, which lists the number of students in the national target population and the number of students excluded at both the school level and within the school for each population to be assessed. An example of a completed Sampling Form 2 is presented in Exhibit 3.2.

Exhibit 3.2: Example of Sampling Form 2

Sampling Form 2 — **Coverage and Exclusions**

See Section 3 of PIRLS 2016 Survey Operations Procedures Unit 1

PIRLS 2016 Participant : < Name of the Country >

1. This Sampling Form refers to:

	PIRLS	PIRLS Literacy
Grade	4	

		Number of schools	Number of students
Total enrollment in the target grade:	[a]	822	56,560

2. School-level exclusions *(if applicable):*

	Description of exclusions	Number of schools	Number of students
1.	Students taught in <language>	8	630
2.	Special education schools	16	325
3.	Very small schools (less than 5 students in grade 4)	40	110
4.			
5.			
TOTAL:	*(Sum of exclusions - Calculated automatically)* [b]	64	1,065

		schools	students
Percentage of school-level exclusions: *(Box [b] ÷ Box [a] x 100)*	**[1]**	7.8%	1.9%
3. Total enrollment after school-level exclusions: *(Box [c] = Box [a] - Box [b])*	[c]	758	55,495

Values calculated automatically

4. Within-school exclusions *(if applicable):*

	Description of exclusions	Number of students
1.	Students with special education needs	640
2.	(based on PIRLS 2011)	
3.		
TOTAL:	*(Sum of exclusions - Calculated automatically)* [d]	640

		schools	students
Expected percentage of within-school exclusions: *(Box [d] ÷ Box [c] x 100)*	**[2]**	0.0%	1.2%
5. **Expected percentage of reduced coverage and exclusions:** *(Box [1] + (1 - Box [1]) X Box [2])*		7.8%	3.0%

Values calculated automatically

6. Total enrollment in the target grade in previous school years.

Years	Number of schools	Number of students
2011/2012	856	58,451
2010/2011	890	61,489

Requirements for Sampling the Target Population

PIRLS sets high standards for sampling precision, participation rates, and sample implementation in order to achieve national samples of the highest quality and survey estimates that are unbiased, accurate, and internationally comparable.

Sampling Precision and Sample Size

Because PIRLS is fundamentally a study of student achievement, the precision of estimates of student achievement is of primary importance. To meet the PIRLS standards for sampling precision, national student samples should provide for a standard error no greater than .035 standard deviation units for the country's mean achievement. This standard error corresponds to a 95% confidence interval of ± 7 score points for the achievement mean and of ± 10 score points for the difference between achievement means from successive cycles (e.g., the difference between a country's achievement mean on PIRLS 2011 and PIRLS 2016)[1]. Sample estimates of any student-level percentage estimate (e.g., a student background characteristic) should have a confidence interval of ± 3.5%.

For most countries, the PIRLS precision requirements are met with a school sample of 150 schools and a student sample of 4,000 students for each target grade. Depending on the average class size in the country, one class from each sampled school may be sufficient to achieve the desired student sample size. For example, if the average class size in a country were 27 students, a single class from each of 150 schools would provide a sample of 4,050 students (assuming full participation by schools and students). Some countries choose to sample more than one class per school, either to increase the size of the student sample or to provide a better estimate of school-level effects.

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.

A school sample larger than the minimum of 150 schools may be required under the following circumstances:

- The average class size in a country is so small that, even when sampling more than one classroom per school, it is not possible to reach the student sample size requirements by selecting only 150 schools.

1 The PIRLS achievement scale was established in 2001 based on the combined achievement distribution of all countries that participated in PIRLS 2001. To provide a point of reference for country comparisons, the scale centerpoint of 500 was located at the mean of the combined achievement distribution. The units of the scale were chosen so that 100 scale score points corresponded to the standard deviation of the distribution.

- Previous cycles of PIRLS showed that the sampling precision requirements cannot be met unless a larger school sample is selected.
- Classes within schools are tracked by student performance. This increases variation between classes in student achievement and can reduce sampling precision. In this situation, it is advisable to sample at least two classrooms per school whenever possible, in addition to sampling more schools.
- A high level of non-response is anticipated, leading to sample attrition and reduced sample size. Note that while a larger school sample helps to maintain sample size in the face of non-response, it does not compensate for non-response bias.

Field Test Sample

The school sample for the PIRLS field test is drawn at the same time and from the same population of schools as the full sample. The field test sample size requirement is 200 students per field test achievement booklet. The total field test sample size is a function of the number of achievement booklets being field tested. Typically, PIRLS has four field test booklets and so requires a field test sample of 800 students. This sample is also used for the online assessment for countries taking part in ePIRLS. For PIRLS 2016, PIRLS Literacy field tested five field test booklets and therefore required a sample size of 1,000 students. As such, countries participating in both PIRLS and PIRLS Literacy required a field test size of 1,800 students.

Participation Rates

To minimize the potential for non-response bias, PIRLS aims for 100% participation by sampled schools, classrooms, and students, while recognizing that some degree of non-participation may be unavoidable. For a national sample to be fully acceptable it must have either:

- A minimum school participation rate of 85%, based on originally sampled schools AND
- A minimum classroom participation rate of 95%, from originally sampled schools and replacement schools AND
- A minimum student participation rate of 85%, from sampled schools and replacement schools

OR

- A minimum combined school, classroom, and student participation rate of 75%, based on originally sampled schools (although classroom and student participation rates may include replacement schools)

Classrooms with less than 50% student participation are deemed to be not participating.

Developing and Implementing the National Sampling Plan

Although National Research Coordinators are responsible for developing and implementing national sampling plans, Statistics Canada and IEA Hamburg work closely with NRCs to help ensure that these sampling plans fully meet the standards set by the TIMSS & PIRLS International Study Center, while also adapting to national circumstances and requirements. National sampling plans must be based on the international two-stage sample design (schools as the first stage and classes within schools as the second stage) and must be approved by Statistics Canada.

PIRLS Stratified Two-Stage Cluster Sample Design

The basic international sample design for PIRLS is a stratified two-stage cluster sample design, as follows:

First Sampling Stage. For the first sampling stage, schools are sampled with probabilities proportional to their size (PPS) from the list of all schools in the population that contain eligible students. The schools in this list (or sampling frame) may be stratified (sorted) according to important demographic variables. Schools for the field test and data collection are sampled simultaneously using a systematic random sampling approach. Two replacement schools are also pre-assigned to each sampled school during the sample selection process, and these replacement schools are held in reserve in case the originally sampled school refuses to participate. Replacement schools are used solely to compensate for sample size losses in the event that the originally sampled school does not participate. School sampling is conducted for each country by Statistics Canada with assistance from IEA Hamburg, using the sampling frame provided by the country's National Research Coordinator.

Second Sampling Stage. The second sampling stage consists of the selection of one (or more) intact class from the target grade of each participating school. Class sampling in each country is conducted by the National Research Coordinator using the Within-School Sampling Software (WinW3S) developed by IEA Hamburg and Statistics Canada. Having secured a sampled school's agreement to participate in the assessment, the National Research Coordinator requests information about the number of classes and teachers in the school and enters it in the WinW3S database. Classes smaller than a specified minimum size are grouped into pseudo-classes prior to sampling. The software selects classes with equal probabilities within schools. All students in each sampled class participate in the assessment. Sampled classes that refuse to participate may not be replaced.

For countries participating in both PIRLS and PIRLS Literacy, students within a sampled class are randomly assigned either a PIRLS or PIRLS Literacy booklet through a booklet rotation system. This is done to ensure that PIRLS and PIRLS Literacy are administered to probabilistically equivalent samples. In countries taking part in ePIRLS, all students assessed in PIRLS are expected to participate in ePIRLS.

Stratification

Stratification consists of arranging the schools in the target population into groups, or strata, that share common characteristics such as geographic region or school type. Examples of stratification variables used in PIRLS include region of the country (e.g., states or provinces); school type or source of funding (e.g., public or private); language of instruction; level of urbanization (e.g., urban or rural area); socioeconomic indicators; and school performance on national examinations.

In PIRLS, stratification is used to:

- Improve the efficiency of the sample design, thereby making survey estimates more reliable
- Apply different sample designs, such as disproportionate sample allocations, to specific groups of schools (e.g., those in certain states or provinces)
- Ensure proportional representation of specific groups of schools in the sample

School stratification can take two forms: explicit and implicit. In explicit stratification, a separate school list or sampling frame is constructed for each stratum and a sample of schools is drawn from that stratum. In PIRLS, the major reason for considering explicit stratification is disproportionate allocation of the school sample across strata. For example, in order to produce equally reliable estimates for each geographic region in a country, explicit stratification by region may be used to ensure the same number of schools in the sample for each region, regardless of the relative population size of the regions.

Implicit stratification consists of sorting the schools by one or more stratification variables within each explicit stratum, or within the entire sampling frame if explicit stratification is not used. The combined use of implicit strata and systematic sampling is a very simple and effective way of ensuring a proportional sample allocation of students across all implicit strata. Implicit stratification also can lead to improved reliability of achievement estimates when the implicit stratification variables are correlated with student achievement.

National Research Coordinators consult with Statistics Canada and IEA Hamburg to identify the stratification variables to be included in their sampling plans. The school sampling frame is sorted by the stratification variables prior to sampling schools so that adjacent schools are as similar as possible. Regardless of any other explicit or implicit variables that may be used, the school size is always included as an implicit stratification variable.

To document the stratification variables used in their sampling plans, each National Research Coordinator completes Sampling Form 3, which lists the variables to be used for explicit and implicit stratification, and the number of levels of each stratification variable. An example of a completed Sampling Form 3 is presented in Exhibit 3.3. Appendix 3A provides the list of explicit and implicit stratification variables used in the sampling process for each the country. Further details on the explicit and implicit stratification variables for each country can be found in the Characteristics of National Samples section in Chapter 5: Sampling Implementation.

Exhibit 3.3: Example of Sampling Form 3

Sampling Form 3	Stratification

See Section 4 of PIRLS 2016 Survey Operations Procedures Unit 1

PIRLS 2016 Participant : < Name of the Country >

1. This Sampling Form refers to:

PIRLS Grade	**PIRLS Literacy** Grade
4	

Stratification of schools

2. List and describe the variables to be used for stratification in order of importance:
(Please note that the choice of variables used for explicit or implicit stratification will be discussed during consultations with Statistics Canada)

Stratification Variables			
	Name	Description	# of levels
1	School type	public, private	2
2	Socio-economic status	high, medium, low	3
3			
4			
5			
6			

Include additional information if necessary:

3. If applicable, describe additional requirements for sub-national estimates (e.g., oversampling of specific groups of the population):

would like to have reliable estimates for students from the private schools

School Sampling Frame

One of the National Research Coordinator's most important sampling tasks is the construction of a school sampling frame for the target population. The sampling frame is a list of all schools in the country that have students enrolled in the target grade, and is the list from which the school sample is drawn. A well-constructed sampling frame provides complete coverage of the national target population without being contaminated by incorrect or duplicate entries or entries that refer to elements that are not part of the defined target population.

A suitable school measure of size (MOS) is a critical aspect of the national sampling plan, because the size of a school determines its probability of selection. The most appropriate school measure of size is an up-to-date count of the number of students in the target grade. If the number of students in the target grade is not available, total student enrollment in the school may be the best available substitute.

Sampling Form 4, presented in Exhibit 3.4, provides some basic information about the school sampling frame, including the average class size at the target grade, the number of classrooms to be sampled per school, the school measure of size (MOS) to be used for school sampling, and the school year from which the frame was constructed.

Exhibit 3.4: Example of Sampling Form 4

Sampling Form 4	Classroom Information and Sampling Frame

See Section 5 of PIRLS 2016 Survey Operations Procedures Unit 1

PIRLS 2016 Participant : < Name of the Country >

1. This Sampling Form refers to:

PIRLS Grade	**PIRLS Literacy** Grade
4	

2. Specify the school measure of size (MOS) to be used.

   Please select the MOS to be used:
   (Click in box and on right arrow to see drop down menu)

   1. Number of students in the target grade (preferred)

   Name of the MOS variable in the school frame: GR4_STD

   If "Other," please describe:

3. Specify the average class size (ACS) for the target grade in your schools. 24

4. Specify how many classrooms you plan to sample per school. *(Click in box and on right arrow to see drop down menu)*

   2. More than one classroom in tracked schools

   If "Other," please describe:

5. Specify the school year for which enrollment data will be used for the school MOS. 2014/2015

6. If a frame other than a single-level sampling frame (list of all schools) is to be used, please provide a preliminary description of the information available to construct this frame.

   n.a.

The school sampling frame is usually a spreadsheet containing a single entry for each school. This entry includes a unique identification number and contact information (if appropriate given the country's privacy laws), the values of the stratification variables for the school, and the school measure of size. It is useful if the school entry also includes the number of classes in the school in the target grade because this provides a mechanism for predicting in advance the size of the eventual student sample. This predicted sample size may be compared with the eventual student sample size as a check on the sampling process.

Exhibit 3.5 provides an example of a partial sampling frame for a country assessing PIRLS 2016. In this example, socioeconomic status and school type could be used as stratification variables.

Exhibit 3.5: Example of a Partial Sampling Frame

	A	B	C	D	E	F	G	H	I	J
1	School ID	SES	Type	Grade 4 Students	Grade 4 Classes	Excluded	School Name	School Address	City	Postal Code
2	E184510	High	Private	177	7	0	Campbell primary School	Rose Ave 86	Red River	104854
3	E136224	Medium	Public	121	5	0	Stromboli Primary School	Abbot's Drive 31	Rosetown	114766
4	E140571	Low	Public	82	3	0	Wooster Elementary School	Woodland Crescent 22	Bloombury	105532
5	E171470	Medium	Public	137	5	0	Alison Primary School	Walter Avenue 99	New Bear City	101425
6	D156972	Medium	Public	234	9	0	Middletown School	Strong Avenue 67	Carrytown	112546
7	E148426	Medium	Public	65	3	0	St John's Primary School	40th Street 13	St John	116044
8	E172294	High	Public	189	8	0	St Mary's Primary School	55th Street 12	Cambelltown	116149
9	X170050	High	Public	150	6	0	Honeydale Primary School	Clarence Road 33	Carlingtown	103153
10	X114207	High	Private	37	1	1	Green Grove Deaf School	Martin Drive 24	Lancaster	107577
11	E199157	High	Private	45	2	0	Stoneborough Primary School	Clare Crescent 7	Blue Lake	103323
12	E171645	High	Public	186	7	0	Pine Tree Primary School	5th Avenue 1	Wallington	118431
13	E128280	Low	Public	32	1	0	Pennington Elementary School	Taylor Ave 25	Wallington	108918
14	X111518	Low	Public	60	2	0	Western Community School	Katryn Road 1	Heaventown	103204
15	X129037	Medium	Private	135	5	0	Central Lane Primary School	Central lane Drive 28	Garden Heights	111495
16	D138959	Low	Public	175	7	0	Tinsdale Elementary School	Kelly Street 18	Holster	106231
17	D155958	High	Public	54	2	0	Douglas Primary School	Alice Street 10	Grahamtown	110874
18	D121060	Low	Private	103	4	0	Dry Creek Primary School	Walnut Drive 25	Grahamtown	106176
19	D143140	Medium	Private	67	3	1	Mad River Specialized School	Maple Crescent 75	Rosendale	118917
20	D152771	Medium	Private	86	3	0	Red Rose Private School	Starlight Avenue 14	Carp	105052
21	D122540	Low	Public	25	1	0	Waterloo Community School	Parkdale Ave 14	Kanata	101670
22	D164184	High	Private	15	1	0	Fruit Tree Primary School	Carling Drive 96	Brookside	118428
23	B141[illegible]			132	5	0	Tinsdale Elementary School	[illegible]	[illegible]rookside	107306
24	[illegible]				3	0	Blueberry Primary School			113976
[illegible]						0	Darlington School			[illegible]8556

Sampling Schools

Once the school sampling frame is structured to meet all international and national requirements, Statistics Canada can draw the school sample. If the sampling frame is explicitly stratified, it is necessary to decide how the school sample is to be allocated among the explicit strata (i.e., the number of schools to be sampled in each stratum). When this has been decided, a sample of schools is selected within each explicit stratum using systematic sampling with probabilities proportional to size. The PPS technique means that the larger schools, those with more students, have a higher probability of being sampled than the smaller schools. However, this difference in the selection probabilities of larger and smaller schools is largely offset at the second stage of sampling by selecting a fixed number of classes (usually one or two) with equal probability from the sampled school. Classes in large schools with many classes at the target grade have a lower probability of selection than classes in smaller schools that have just one or two classes. A description of the school sampling procedure is provided in Appendix 3B.

Even though the field test is scheduled in the school year before the year of data collection in most countries, the preferred approach in PIRLS is to select both samples of schools at the same time. This ensures that both the field test and data collection samples constitute random samples representative of all schools in the country, and that no school is selected for both samples.

Replacement Schools. Ideally, all schools sampled for PIRLS should participate in the assessments, and National Research Coordinators work hard to achieve this goal. Nevertheless, it is anticipated that a 100 percent participation rate may not be possible in all countries. To avoid sample size losses, the sampling plan identifies, *a priori*, specific replacement schools for each sampled school. Each originally sampled school has two pre-assigned replacement schools, usually the school immediately preceding the originally sampled school on the school sampling frame and the one immediately following it. Replacement schools always belong to the same explicit stratum as the original but may come from different implicit strata if the school they are replacing is either the first or last school of an implicit stratum.

The main justification for replacement schools in PIRLS is to ensure adequate sample sizes for analysis of subpopulation differences. Although the use of replacement schools does not eliminate the risk of bias due to school nonparticipation, employing implicit stratification and ordering the school sampling frame by school size increases the chances that a sampled school's replacements would have similar characteristics. This approach maintains the desired sample size while restricting replacement schools to strata where nonresponse occurs. Since the school frame is ordered by school size, replacement schools also tend to be similar in size to the school they are designated to replace.

National Research Coordinators understand that they should make every effort to secure the participation of all of the sampled schools. Only after all attempts to persuade a sampled school to participate have failed is the use of its replacement school considered.

Common Adjustments to the PIRLS School Sampling Design

The PIRLS school sample design offers considerable flexibility in allowing countries to control the overlap with other national or international assessments. In some cases, countries try to ensure that assessments are spread across schools and therefore prefer that PIRLS sampling avoid, when possible, selecting schools that have recently administered other national and international assessments. To provide flexibility to meet these requests, Statistics Canada implements modified sampling procedures—the details of which are described in Appendix 3C.

Sampling Classes

Within each sampled school, all classes with students at the target grade are listed, and one or more intact classes are selected with equal probability of selection using systematic random sampling. This procedure is implemented using the WinW3S sampling software. The selection of classes with equal probability, combined with the PPS sampling method for schools, in general results in a self-weighting student sample. If the school has multi-grade classes (i.e., the class contains students from more than one grade level), only students from the target grade are eligible for sampling.

When a country participates in both PIRLS and PIRLS Literacy, students within the sampled classes are randomly assigned to one study or the other by rotating the PIRLS and PIRLS Literacy booklets within the sampled classes. This is done automatically by the WinW3S software.

Because small classes tend to increase the risk of unreliable survey estimates and can lead to reduced overall student sample size, it is necessary to avoid sampling too many small classes. Based on consideration of the size distribution of classes and the average class size, a lower class size limit or minimum class size (MCS) is specified for each country. Prior to sampling classes in a school, any class smaller than half the MCS is combined with another class in the school to form a "pseudo-class" for sampling purposes. The procedure for sampling classes within schools is described in more detail in the Survey Operations Procedures chapter of this publication.

Sampling Weights

National student samples in PIRLS are designed to accurately represent the target population within a specified margin of sampling error, as described previously. After the data have been collected and processed, sample statistics such as means and percentages that describe student characteristics are computed as weighted estimates of the corresponding population parameters, where the weighting factor is the sampling weight. A student's sampling weight is essentially the inverse of the student's probability of selection, with appropriate adjustments for nonresponse. In principle, the stratified two-stage sampling procedure used in PIRLS, where schools are sampled with probability proportional to school size and classes are sampled with probability inversely proportional to school size, provides student samples with equal selection probabilities. However, in practice disproportionate sampling across explicit strata by varying the number of classes selected and differential patterns of nonresponse can result in varying selection probabilities, requiring a unique sampling weight for the students in each participating class in the study.

The student sampling weight in PIRLS is a combination of weighting components reflecting selection probabilities and sampling outcomes at three levels—school, class, and student. At each level, the weighting component consists of a basic weight that is the inverse of the probability of selection at that level, together with an adjustment for nonparticipation. The overall sampling weight for each student is the product of the three weighting components: school, class (within school), and student (within class).

For countries participating in both PIRLS and PIRLS Literacy, sampling weights are calculated independently for each study. Although all students participating in PIRLS were also supposed to participate in ePIRLS, in practice this was always the case either by design (e.g., subsampling of schools, classes, or students), or by circumstance, (i.e., student absences). Consequently, the ePIRLS samples were also weighted separately. Further details on the special weight adjustments for ePIRLS can be found in Chapter 5: Sampling Implementation for PIRLS.

School Weighting Component

Given that schools in PIRLS are sampled with probability proportional to school size, the basic school weight for the i^{th} sampled school (i.e., the inverse of the probability of the ith school being sampled) is defined as:

$$BW_{sc}^{i} = \frac{M}{n \cdot m_i} \tag{3.1}$$

where n is the number of sampled schools, m_i is the measure of size for the i^{th} school, and

$$M = \sum_{i=1}^{N} m_i \tag{3.2}$$

where N is the total number of schools in the explicit stratum.[2]

School Nonparticipation Adjustment. If a sampled school does not participate in PIRLS and its two designated replacement schools do not participate, it is necessary to adjust the basic school weight to compensate for the reduction in sample size. The school-level nonparticipation adjustment is calculated separately for each explicit stratum, as follows:

$$A_{sc} = \frac{n_s + n_{r1} + n_{r2} + n_{nr}}{n_s + n_{r1} + n_{r2}} \tag{3.3}$$

where n_s is the number of originally sampled schools that participated, n_{r1} and n_{r2} the number of first and second replacement schools, respectively, that participated, and n_{nr} is the number of schools that did not participate. Sampled schools that are found to be ineligible[3] are not included in the calculation of this adjustment.

Combining the basic school weight and the school nonparticipation adjustment, the final school weighting component for the i^{th} school becomes:

$$FW_{sc}^{i} = A_{sc} \cdot BW_{sc}^{i} \tag{3.4}$$

2 For countries such as the Russian Federation that include a preliminary sampling stage, the basic school weight also incorporates the probability of selection in this preliminary stage. The basic school weight in such cases is the product of the preliminary stage weight and the school weight.

3 A sampled school is ineligible if it is found to contain no eligible students (i.e., no students in the target grade). Such schools usually are in the sampling frame by mistake or are schools that recently have closed.

It should be noted that, as well as being a crucial component of the overall student weight, the final school weighting component is a sampling weight in its own right, and can be used in analyses where the school is the unit of analysis.

Class Weighting Component

The class weighting component reflects the class-within-school selection probability. After a school has been sampled and has agreed to participate in PIRLS, one or two classes are sampled with equal probability from the list of all classes in the school at the target grade. Because larger schools have more classes from which to sample than smaller schools, the probability of class selection varies with school size, with students in small schools more likely to have their class selected than students in large schools. This relatively greater selection probability for students in small schools offsets their lower selection probability at the first stage, where probability-proportional-to-size school sampling results in higher selection probabilities for larger schools.

The basic class-within-school weight for a sampled class is the inverse of the probability of the class being selected from all of the classes in its school. For the i^{th} sampled school, let C^i be the total number of eligible classes and c^i the number of sampled classes. Using equal probability sampling, the basic class weight for all sampled classes in the i^{th} school is:

$$BW_{cl}^{i} = \frac{C^{i}}{c^{i}} \tag{3.5}$$

For most PIRLS participants, c^i takes the values 1 or 2.

Class Nonparticipation Adjustment. Basic class weights are calculated for all sampled classes in the sampled and replacement schools that participate in PIRLS. A class-level nonparticipation adjustment is applied to compensate for classes that do not participate or where the student participation rate is below 50 percent.[4] Such sampled classes are assigned a weight of zero. Class nonparticipation adjustments are applied at the explicit stratum level rather than at the school level to minimize the risk of bias. The adjustment is calculated as follows:

$$A_{cl} = \frac{\sum_{i}^{s+r1+r2} 1}{\sum_{i}^{s+r1+r2} \delta_i / c^i} \tag{3.6}$$

where c^i is the number of sampled classes in the i^{th} school, as defined earlier, and δ_i gives the number of participating classes in the i^{th} school.

4 Although sampling weights are calculated separately for each study when countries participate in both PIRLS and PIRLS Literacy, the criteria to evaluate if student participation within a class is below 50% uses the student participation from both studies combined. Therefore, if 50% or more students from a class participated in either PIRLS or PIRLS Literacy, the class is considered as participating when calculating sampling weights for PIRLS or PIRLS Literacy.

Combining the basic class weight and the class nonparticipation adjustment, the final class weighting component, assigned to all sampled classes in the i^{th} school, becomes:

$$FW_{cl}^{i,j} = A_{cl} \cdot BW_{cl}^{i} \tag{3.7}$$

Student Weighting Component

The student weighting component represents the student-within-class selection probability. The basic student weight is the inverse of the probability of a student in a sampled class being selected.

In the typical PIRLS situation where intact classes are sampled, all students in the class are included, and so this probability is unity. However, under certain circumstances, students may be sampled within the class, and in these circumstances the probability is less than unity. For PIRLS 2016, within-class sampling occurred in countries that decided to administer both PIRLS and PIRLS Literacy.

For an intact class with no student subsampling, the basic student weight for the j^{th} class in the i^{th} school is computed as follows:

$$BW_{st1}^{i,j} = 1.0 \tag{3.8}$$

For classes with student subsampling, the basic student weight for the j^{th} class in the i^{th} school is:

$$BW_{st2}^{ij} = \frac{n_{rg}^{i,j} + n_{bs}^{i,j}}{n_{rg}^{i,j}} \tag{3.9}$$

where $n_{rg}^{i,j}$ is the number of students in the j^{th} class of the i^{th} school selected to participate in PIRLS and $n_{bs}^{i,j}$ is the number of students in the class not selected.[5] In the case of countries administering both PIRLS and PIRLS Literacy, a set of weights is calculated for each study and the basic student weight is calculated differently, as the participation status is known for all the students in each sampled class. In this case, the basic student weight for the j^{th} class in the i^{th} school for study k is given by:

$$BW_{st3}^{ij} = \begin{cases} 1 & \text{for students who left school or were excluded,} \\[2ex] \dfrac{n_{rg'}^{i,j} + n_{bs'}^{i,j}}{n_{rg'}^{i,j}} & \text{for all other students selected for study } k \end{cases} \tag{3.10}$$

5 In one ePIRLS country with limited access to computers in school, a random subsampling mechanism was put in place to subsample students within class for the ePIRLS assessment, resulting in student sampling probability less than unity for ePIRLS.

where k represents either PIRLS or PIRLS Literacy, $n_{rg'}^{i,j}$ and $n_{bs'}^{i,j}$ represent the number of students in the j^{th} class of the i^{th} school selected to participate in study k and the number of students in the j^{th} class of the i^{th} school not selected for study k respectively, without counting students who either were excluded or left school after the class listing was completed.

Adjustment for Non-Participation. The student nonparticipation adjustment for the j^{th} classroom in the i^{th} school is calculated as:

$$A_{st1}^{i,j} = A_{st2}^{i,j} = A_{st3}^{i,j} = \frac{s_{rs}^{i,j} + s_{nr}^{i,j}}{s_{rs}^{i,j}} \tag{3.11}$$

where $s_{rs}^{i,j}$ is the number of participating students (i.e., students that participated in PIRLS or PIRLS Literacy and have assessment scores) in the j^{th} class of the i^{th} school and $s_{nr}^{i,j}$ is the number of students sampled in this class who were expected to have assessment scores but did not participate in the assessment. For intact classes, the sum of $s_{rs}^{i,j}$ and $s_{nr}^{i,j}$ is the total number of students listed in the class, not counting excluded students or students who have left the school since class list was published.

The final student weighting component for students in the j^{th} classroom of the i^{th} school is:

$$FW_{st}^{i,j} = A_{st\Delta}^{i,j} \cdot BW_{st\Delta}^{i,j} \tag{3.12}$$

where Δ equals 1 when there was no student subsampling (intact classes), 2 when a sample of students was drawn from the students in the class, and 3 when both PIRLS and PIRLS Literacy were administered within the same schools and classes.

Overall Student Sampling Weight. The overall student sampling weight is the product of the final weighting components for schools, classes, and students, as follows:

$$W^{i,j} = FW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j} \tag{3.13}$$

Overall student sampling weights are only attributed to participating students, with non-participants weighted at 0. All student data reported in the PIRLS international reports are weighted by the overall student sampling weight, known as TOTWGT in the PIRLS international databases.

Participation Rates

Because nonparticipation can result in sample bias and misleading results, it is important that the schools, classes, and students that are sampled to participate in PIRLS actually take part in the assessments. To show the level of sampling participation in each country, PIRLS calculates both unweighted participation rates (i.e., based on simple counts of schools, classes, and students) and weighted participation rates based on the sampling weights described in the previous section. Unweighted participation rates provide a preliminary indicator that may be used to monitor progress in securing the participation of schools and classes, whereas weighted participation rates are the ultimate measure of sampling participation.

PIRLS reports weighted participation rates as well as unweighted participation rates for schools, classes, and students, and overall participation rates that are a combination of all three. To distinguish between participation based solely on originally sampled schools and participation that also relies on replacement schools, school and overall participation rates are computed separately for originally sampled schools only and for originally sampled together with replacement schools.

Unweighted School Participation Rate

The unweighted school participation rate is the ratio of the number of participating schools to the number of originally sampled schools, excluding any sampled schools found to be ineligible. A school is considered to be a "participating school" if at least one of its sampled classes has a student participation rate of at least 50 percent. The two unweighted school participation rates are calculated as follows:

R_{unw}^{sc-s} = unweighted school participation rate for originally sampled schools only

R_{unw}^{sc-r} = unweighted school participation rate, including originally sampled and first and second replacement schools

$$R_{unw}^{sc-s} = \frac{n_s}{n_s + n_{r1} + n_{r2} + n_{nr}} \tag{3.14}$$

$$R_{unw}^{sc-r} = \frac{n_s + n_{r1} + n_{r2}}{n_s + n_{r1} + n_{r2} + n_{nr}} \tag{3.15}$$

Unweighted Class Participation Rate

The unweighted class participation rate is the ratio of the number of sampled classes that participated to the number of classes sampled, as follows:

$$R_{unw}^{cl} = \frac{\sum_{i}^{s+r1+r2} c_*^i}{\sum_{i}^{s+r1+r2} c^i} \tag{3.16}$$

where c^i is the number of sampled classes in the i^{th} school, and c_*^i is the number of participating classes in the i^{th} school. Both summations are across all participating schools.

Unweighted Student Participation Rate

The unweighted student participation rate is the ratio of the number of selected students that participated in PIRLS to the total number of selected students that should have been assessed in the participating schools and classes. Classes where less than 50 percent of the students participate are considered to be not participating, and so students in such classes also are considered to be nonparticipants.[6] The unweighted student participation rate is computed as follows:

$$R_{unw}^{st} = \frac{\sum_{i,j} s_{rs}^{i,j}}{\sum_{i,j} s_{rs}^{i,j} + \sum_{i,j} s_{nr}^{i,j}} \tag{3.17}$$

Overall Unweighted Participation Rate

The overall unweighted participation rate is the product of the unweighted school, class, and student participation rates. Because PIRLS computes two versions of the unweighted school participation rate, one based on originally sampled schools only and the other including replacements as well as originally sampled schools, there also are two overall unweighted participation rates:

R_{unw}^{ov-s} = unweighted overall participation rate for originally sampled schools only

R_{unw}^{ov-r} = unweighted overall participation rate, including originally sampled and first and second replacement schools

$$R_{unw}^{ov-s} = R_{unw}^{sc-s} \cdot R_{unw}^{cl} \cdot R_{unw}^{st} \tag{3.18}$$

$$R_{unw}^{ov-r} = R_{unw}^{sc-r} \cdot R_{unw}^{cl} \cdot R_{unw}^{st} \tag{3.19}$$

Weighted School Participation Rate

The weighted school participation rate is the ratio of two estimates of the size of the target student population. The numerator is derived from the measure of size of those sampled schools that participated in PIRLS and the denominator is the weighted estimate of the total student enrollment in the population. Weighted school participation rates are computed for originally sampled schools and for originally sampled and replacement schools combined, as follows:

6 For countries that participated in both PIRLS and PIRLS Literacy, this 50% criteria is applied to student participation from both studies combined.

R_{wtd}^{sc-s} = weighted school participation rate for originally sampled schools only

R_{wtd}^{sc-r} = weighted school participation rate, including originally sampled and first and second replacement schools

$$R_{wtd}^{sc-s} = \frac{\sum_{i,j}^{s} BW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j}}{\sum_{i,j}^{s+r1+r2} FW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j}} \tag{3.20}$$

$$R_{wtd}^{sc-r} = \frac{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j}}{\sum_{i,j}^{s+r1+r2} FW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j}} \tag{3.21}$$

Summations in both the numerator and denominator are over all responding students and include appropriate class and student sampling weights. Note that the basic school weight appears in the numerator, whereas the final school weight appears in the denominator.

Weighted Class Participation Rate

The weighted class participation rate is computed as follows:

$$R_{wtd}^{cl} = \frac{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot BW_{cl}^{i,j} \cdot FW_{st}^{i,j}}{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot FW_{cl}^{i,j} \cdot FW_{st}^{i,j}} \tag{3.22}$$

where both the numerator and denominator are summations over all responding students from classes with at least 50 percent of their students participating in the study, and the appropriate student-level sampling weights are used. In this formula, the basic class weight appears in the numerator, whereas the final class weight appears in the denominator. And, the denominator in this formula is the same quantity that appears in the numerator of the weighted school participation rate for all schools, whether originally sampled or replacement.

Weighted Student Participation Rate

The weighted student participation rate is computed as follows:

$$R_{wtd}^{st} = \frac{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot BW_{cl}^{i,j} \cdot BW_{st}^{i,j}}{\sum_{i,j}^{s+r1+r2} BW_{sc}^{i} \cdot BW_{cl}^{i,j} \cdot FW_{st}^{i,j}} \tag{3.23}$$

where both the numerator and denominator are summations over all responding students from participating schools. In this formula, the basic student weight appears in the numerator, whereas the final student weight appears in the denominator. Also, the denominator in this formula is the same quantity that appears in the numerator of the weighted class participation rate for all participating schools, whether originally sampled or replacement.

Overall Weighted Participation Rate

The overall weighted participation rate is the product of the weighted school, class, and student participation rates. Because there are two versions of the weighted school participation rate, one based on originally sampled schools only and the other including replacement as well as originally sampled schools, there also are two overall weighted participation rates:

R_{wtd}^{ov-s} = weighted overall participation rate for originally sampled schools only

R_{wtd}^{ov-r} = weighted overall participation rate, including sampled, first and second replacement schools

$$R_{wtd}^{ov-s} = R_{wtd}^{ov-s} \cdot R_{wtd}^{cl} \cdot R_{wtd}^{st} \tag{3.24}$$

$$R_{wtd}^{ov-r} = R_{wtd}^{sc-r} \cdot R_{wtd}^{cl} \cdot R_{wtd}^{st} \tag{3.25}$$

Weighted school, class, student, and overall participation rates are computed for each PIRLS participant using these procedures.

References

Chowdhury, S., Chu, A., & Kaufman, S. (2000). Minimizing overlap in NCES surveys. *Proceedings of the Survey Methods Research Section, American Statistical Association*, 174-179. Retrieved from http://www.amstat.org/sections/srms/Proceedings/papers/2000_025.pdf

UNESCO. (2012). *International Standard Classification of Education ISCED 2011.* Montreal: UNESCO Institute of Statistics. Retrieved from http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf

Appendix 3A: PIRLS 2016 Stratification Variables

PIRLS 2016 Stratification Variables

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Australia	State or territory (8)	8	Geographic location (3) School type (3) Socioeconomic status (2)
Austria	Region (9)	9	None
Azerbaijan	Language (2) Urbanization (2) City (2)	4	None
Bahrain	Governorate (5) Gender (2)	9	None
Belgium (Flemish)	Region (6) School type (3) Socioeconomic status (4)	18	None
Belgium (French)	School network (3) Socioeconomic status (4)	10	None
Bulgaria	School type (3) Urbanization (3)	8	Urbanization (2)
Canada	Province (8) School language (2) within British Columbia School type (3) within English schools in British Columbia School system (2) within Alberta School type (2) within Alberta School type (3) within Ontario (3) Language (2) within Ontario School language (2) within Quebec School type (2) within Quebec School language (2) within New Brunswick	22	Region (4) in public and Catholic schools within Ontario
Chile	School type (3) Urbanization (2) School size (3)	8	None
Chinese Taipei	Region (4)	4	None
Czech Republic	Region (14)	14	None
Denmark	School type (2)	2	None
Egypt	Region (3) School type (2)	6	Urbanization (2)
England	School type (2) Performance (5)	6	None
Finland	Language (2) Major region (4) Urbanization (2)	8	None
France	School type (3)	3	None

PIRLS 2016 Stratification Variables (Continued)

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Georgia	Language taught in school (2) Teacher certification (2) Urbanization (2) School type (2)	7	None
Germany	Federal state (5)	5	Percentage of immigrants (4)
Hong Kong SAR	School gender (2) School type (4)	5	None
Hungary	Community type (3) National assessment reading score (4)	11	None
Iran, Islamic Rep. of	School type (2) Gender (3) Region group (3) Province (6)	22	None
Ireland	School-level socioeconomic status (4) Language (3) Gender (3)	8	None
Israel	School sector (3) Socioeconomic status (3) Subgroups within Arab sector (3)	10	None
Italy	School type (2) Region (5)	6	None
Kazakhstan	Region (4) Language (3) Urbanization (2)	17	None
Kuwait	School type (2) Region (6) Gender (2) Language (3)	15	None
Latvia	School level (2) Urbanization (3) Language (2) School type (2)	10	None
Lithuania	Language (4) Urbanization (4)	7	None
Macao SAR	None	1	None
Malta	School type (3)	3	None
Morocco	School type (2) Region (16)	18	None
Netherlands	Socioeconomic status level (5) Urbanization (5)	12	None
New Zealand	School type (4) Socioeconomic status (4) Urbanization (2)	11	None
Northern Ireland	Region (5) Deprivation (5)	14	None

PIRLS 2016 Stratification Variables (Continued)

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Norway (5)	Grade 5 only/grade 4 and 5 schools (2) Language (2)	3	None
Oman	Governorates (11) School type (3)	13	None
Poland	Urbanization (4) School performance level (5)	15	None
Portugal	Region (7) School type (2)	8	None
Qatar	School type (3) Gender (3)	4	None
Russian Federation	Region (42)	42	None
Saudi Arabia	Region (13) Gender (2)	17	None
Singapore	None	1	None
Slovak Republic	Language (2) Socioeconomic status (3) Grouped region (5)	14	None
Slovenia	School type (2) Region (4)	8	None
South Africa (5)	Language (11) Province (9)	23	None
Spain	Region (7) School type (2) Bilingual status (2) within Madrid strata	19	None
Sweden	Grade average (4)	4	None
Trinidad and Tobago	School type (2) Region (8)	9	None
United Arab Emirates	Emirates (7) School type (2) Language of instruction (2) Region (3) within Abu Dhabi School type (3) within Abu Dhabi Curriculum (4) within Abu Dhabi School type (2) within Dubai Language (3) within Dubai	28	None
United States	Poverty level (2) School type (2) Census region (4)	12	Urbanization (4) Ethnicity Status (2)

PIRLS 2016 Stratification Variables (Continued)

Country	Explicit Stratification Variables	Number of Explicit Strata	Implicit Stratification Variables
Benchmarking Participants			
Andalusia, Spain	School type (2)	2	None
Denmark (3)	School type (2)	2	None
Eng/Afr/Zulu - RSA (5)	Language (3)	5	None
Madrid, Spain	School type (3) Bilingual status (2) within Madrid strata	5	None
Moscow City, Russian Fed.	None	1	None
Buenos Aires, Argentina	School type (2) Socioeconomic status (3)	6	None
Ontario, Canada	School type (3) Language (2)	4	Region (4) in public and Catholic
Quebec, Canada	Language (2) School type (2)	4	None
Norway (4)	Grade 4 only/Grade 4 and 5 schools (2) Language (2)	3	None
Abu Dhabi, UAE	Region (3) School type (3) Curriculum (4)	9	None
Dubai, UAE	School type (2) Language (3)	4	None

Appendix 3B: Sampling Schools

PIRLS employs random-start fixed-interval systematic sampling to draw the school sample, with each school selected with probability proportional to its size (PPS).

To sample schools using the PPS systematic sampling method, the schools from each explicit stratum in the sampling frame are sorted by implicit stratification variables and by their measure of size (MOS), as shown in the example. The MOS is accumulated from school to school and the running total (the Cumulative MOS) is listed next to each school. The cumulative MOS across the entire stratum (the Total Measure of Size) is a measure of the size of the school population in the stratum (59,614 students in the example).

First Step: Compute the Sampling Interval

Dividing the Total MOS by the number of schools required for the sample (50 in the example) gives the sampling interval.

- **59,614 ÷ 50 = 1,192.2800**

Second Step: Generate a Random Start

Generate a random number from a uniform (0,1) distribution and multiply it by the sampling interval. The school whose cumulative MOS contains the resulting number is the first school in the sample.

- **0.5481 × 1,192.2800 = 653.4887**
- **School 1718**, with cumulative **MOS** of **690**, is the first school in the sample.

Third Step: Identify the Next School in the Sample (repeat until all schools have been sampled)

- Add the sampling interval to the number computed in the previous step.
- **653.4887 + 1,192.2800 = 1,845.7687**
- **School 0067,** with cumulative **MOS** of **1,855**, is the second school in the sample.
- Repeat until all schools have been sampled. For example, to identify the third school:
- **1,845.7687 + 1,192.2800 = 3,038.0487**
- **School 0333**, with cumulative **MOS** of **3,038**, is the third school in the sample.

Fourth Step: Identify Replacement Schools

Two replacement schools are identified for each sampled school. The first replacement (R1) is the school that immediately follows the sampled school in the sampling frame, and the second replacement (R2) the school that immediately precedes the sampled school.

PPS Systemic Sampling—Schools

Sampling Parameters	
Total Number of schools:	2,119
Total Measure of Size:	59,614
School Sample Size:	50
Sampling Interval:	1,192.2800
Random Start:	653.4887
First Step	
Compute the Sampling Interval:	
59,6914 ÷ 50 = 1,192.2800	
Second Step	
Generate a random start:	
0.5481 x 1,192.2800 = 653.4887	
Third Step (repeat until complete)	
Compute the next selection numbers:	
653.4887 + 1,192.2800 = 1,845.7687	
1,845.7687 + 1,192.2800 = 3,038.0487	
Fourth Step	
Identify Replacement Schools	
(R1, R2)	

School Identifier	School MOS	Cumulative MOS	Sampled Schools
0829	110	110	
0552	101	211	
1802	98	309	
1288	98	407	
2043	95	502	
0974	94	596	***R2***
1718	94	690	✓
1807	93	783	***R1***
0457	93	876	
0244	93	969	
1817	91	1,060	
1741	90	1,150	
1652	89	1,239	
0121	89	1,328	
0309	89	1,417	
0032	89	1,506	
0021	89	1,595	
0609	88	1,683	
0399	86	1,769	***R2***
0067	86	1,855	✓
0202	86	1,941	***R1***
0063	86	2,027	
1467	86	2,113	
1381	86	2,199	
1043	84	2,283	
1318	84	2,367	
0659	84	2,451	
0612	83	2,534	
1696	82	2,616	
0867	82	2,698	
0537	81	2,779	
1794	80	2,859	
0695	80	2,939	
0031	80	3.019	R2
0333	79	3,098	✓
0051	79	3,177	R1
0384	79	3,256	
1361	79	3,335	
1189	79	3,414	
0731	78	3,492	
0634	78	3,570	
1230	77	3,647	

Appendix 3C: School Sampling Design Options to Accommodate Other Samples

PIRLS provides an optional modification to its sampling design for countries that want to minimize overlap between schools sampled for PIRLS and schools sampled for other national or international assessments.

The special sampling procedure implemented by Statistics Canada used a technique described in Chowdhury, Chu, and Kaufman (2000). As explained by the authors, the method can be used to either minimize or maximize overlap amongst several samples. This method is illustrated below with an example where the aim was to minimize the overlap between a current sample of schools S_2 and a previously selected school sample S_1. (For a complete description of the method, readers are referred to the original paper).

Let RL (Response Load) be the number of times a school was sampled from previous samples. In this example, given that there is only one previous sample, RL takes the value 1 if the school was already selected and 0 otherwise.

Given that the RL variable splits the current school frame in two distinct subsets of schools, S_1 where RL=1 and $\overline{S}_1$ where RL=0, we have the following relation:

$$P_i(S_2) = P_i(S_2|S_1) \cdot P_i(S_1) + P_i(S_2|\overline{S}_1) \cdot P_i(\overline{S}_1) \tag{3.26}$$

where $P_i(S_j)$ gives the probability that school i be selected in the sample (S_j), and $P_i(S_j|S_k)$ gives the probability that school i be selected in sample (S_j) given that school i already belongs to (S_k). The idea here is to derive the conditional probabilities in such a way that the unconditional probability of selecting a school in the current sample, $P_i(S_2)$, be equal to the expected probability (as defined by the PIRLS sample design).

Note that the first term after the equal sign in Equation (3.26) is related to cases where the school response load is one, while the last term is related to cases where the school response load is zero. Therefore, minimizing the sample overlap is equivalent to zeroing the first term. In such case, Equation (3.26) becomes:

$$P_i(S_2) = 0 \cdot P_i(S_1) + P_i(S_2|\overline{S}_1) \cdot P_i(\overline{S}_1) \tag{3.27}$$

and consequently,

$$P_i(S_2|\overline{S}_1) = P_i(S_2)/P_i(\overline{S}_1) \tag{3.28}$$

In other words, in the current sample S_2, schools would be selected with the following conditional probabilities:

$$\begin{cases} 0 & \text{if school } i \text{ was already selected in the first sample,} \\ P_i(S_2)/P_i(\overline{S}_1) & \text{otherwise} \end{cases} \tag{3.29}$$

However, Equation (3.26) no longer holds if expression $P_i(S_2)/P_i(\overline{S}_1)$ is greater than one. This can be avoided by setting one as an upper bound. We now have the following expression:

$$P_i(S_2) = P_i(S_2|S_1) \cdot P_i(S_1) + 1 \cdot P_i(\overline{S}_1) \tag{3.30}$$

and consequently

$$\frac{P_i(S_2) - P_i(\overline{S}_1)}{P_i(S_1)} = P_i(S_2|S_1) \tag{3.31}$$

Combining these two results, the conditional probabilities to use when selecting the current sample of schools are given by:

$$\begin{cases} \text{Max} \left[0\,, \quad \dfrac{P_i(S_2) - P_i(\overline{S}_1)}{P_i(S_1)} \right] & \text{if school } i \text{ was already selected in the first sample,} \\ \text{Min} \left[\dfrac{P_i(S_2)}{P_i(\overline{S}_1)}\,, \quad 1 \right] & \text{otherwise} \end{cases} \tag{3.32}$$

Note that maximizing rather than minimizing the overlap between two studies can be done by simply zeroing the last term of Equation (3.26) rather than zeroing the first term, and following the above logic to get the conditional probabilities. The Chowdhury, Chu, and Kaufman (2000) method can be generalized to more than two samples as described in their paper.

Further details about the implementation of this method for the countries and benchmark participants can be found in Chapter 5.

CHAPTER 4

Estimating Standard Errors in the PIRLS 2016 Results

Pierre Foy
Sylvie LaRoche

To obtain estimates of students' proficiency in reading that are both accurate and cost-effective, PIRLS 2016 made extensive use of probability sampling techniques to sample students from the national fourth grade student population, and applied matrix-sampling assessment designs to target individual students with a subset of the complete pool of assessment items. This approach made efficient use of resources, in particular keeping student response burden to a minimum, but at a cost of some variance or uncertainty in the reported statistics, such as the means and percentages computed to estimate population parameters.

To quantify this uncertainty, each statistic in the PIRLS 2016 and ePIRLS 2016 international reports is accompanied by an estimate of its standard error. For statistics reporting student achievement, which are based on plausible values, standard errors have two components. The first reflects the uncertainty due to generalizing from student samples to the entire fourth grade student population, referred to as sampling variance, and the second reflects uncertainty due to inferring students' performance on the entire assessment from their performance on the subset of items that they took, known as imputation variance. For parameter estimates of variables that are not plausible values, standard errors are based entirely on sampling variance.

Estimating Sampling Variance

PIRLS makes extensive use of probability sampling to derive achievement results from national samples of students. Because many such samples are possible but only one sample is drawn, some uncertainty about how well the sample represents the population is to be expected. The uncertainty caused by sampling students from a target population, known as sampling variance, can be estimated from the data of the one sample drawn.

Whereas estimating the sampling variance from simple random samples is a relatively easy task, estimating the sampling variance from the complex sample design of PIRLS is a more challenging endeavor.

A common way to estimate the sampling variance in multistage cluster sampling designs is through resampling schemes such as the balanced repeated replication and Jackknife techniques (Johnson & Rust, 1992; Wolter, 1985). PIRLS uses one variation of the Jackknife, the Jackknife Repeated Replication (JRR), to estimate sampling variances. JRR was chosen because it is computationally straightforward and provides approximately unbiased estimates of the sampling variances and sampling errors of means, total, and percentages.

At the core of the JRR technique is the grouping of primary sampling units—schools—into zones based on the sample design stratification and subsequent repeated draws of subsamples from these zones, i.e., repeated replication. For PIRLS, the two main features of the PIRLS sample design that JRR incorporates in its repeated draws of subsamples are the stratification of schools and the clustering of students within schools. This is done by defining Jackknife sampling zones according to the stratification scheme, pairing successive schools[1] to model the clustering from each national sample (see Chapter 3 for information on the Sample Design). Since most national samples consist of 150 schools, a total of 75 zones are created. If more than 150 schools are selected, then the additional zones are collapsed into the first 75 zones. The subsampling required by JRR is applied within each sampling zone.

Sampling zones are constructed within explicit strata. When an explicit stratum has an odd number of schools, either by design or because of school non-response, the students in the remaining school are randomly divided to make up two "quasi" schools.[2] Each sampling zone then consists of a pair of schools or "quasi" schools.

Exhibit 4.1 lists the number of sampling zones for each PIRLS and ePIRLS 2016 participating country.

1 When schools are sampled, schools are ordered within explicit strata by implicit stratification variables and the measure of size. Based on this sorting, successively sampled schools are matched and classified together in each sampling zone. More information can be found in Appendix 3A of Chapter 3.

2 For example, if a remaining school consists of 2 sampled classrooms, each classroom becomes a "quasi" school.

Exhibit 4.1: Number of Sampling Zones for Each PIRLS 2016 Participating Country

Country	PIRLS and ePIRLS 2016 Sampling Zones	
	PIRLS	ePIRLS
Australia	75	-
Austria	75	-
Azerbaijan	75	-
Bahrain	75	-
Belgium (Flemish)	75	-
Belgium (French)	75	-
Bulgaria	75	-
Canada	75	75
Chile	75	-
Chinese Taipei	75	75
Czech Republic	75	-
Denmark	75	72
Egypt	75	-
England	75	-
Finland	75	-
France	75	-
Georgia	75	75
Germany	75	-
Hong Kong SAR	70	-
Hungary	75	-
Iran, Islamic Rep. of	75	-
Ireland	75	74
Israel	75	75
Italy	75	75
Kazakhstan	75	-
Kuwait	75	-
Latvia	74	-
Lithuania	75	-
Macao SAR	75	-
Malta	75	-
Morocco	75	-
Netherlands	68	-
New Zealand	75	-
Northern Ireland	70	-
Norway (5)	75	71
Oman	75	-

Exhibit 4.1: Number of Sampling Zones for Each PIRLS 2016 Participating Country (Continued)

Country	PIRLS and ePIRLS 2016 Sampling Zones	
	PIRLS	ePIRLS
Poland	75	-
Portugal	75	75
Qatar	75	-
Russian Federation	75	-
Saudi Arabia	75	-
Singapore	75	75
Slovak Republic	75	-
Slovenia	75	75
South Africa	75	-
Spain	75	-
Sweden	75	73
Trinidad and Tobago	75	-
United Arab Emirates	75	75
United States	75	75
Benchmarking Participants		
Buenos Aires, Argentina	75	-
Ontario, Canada	75	-
Quebec, Canada	65	-
Denmark (3)	75	-
Norway (4)	75	-
Moscow City, Russian Fed.	75	-
Eng/Afr/Zulu - RSA (5)	64	-
Andalusia, Spain	75	-
Madrid, Spain	75	-
Abu Dhabi, UAE	72	73
Dubai, UAE	75	75

The JRR procedure draws two subsamples from each sampling zone: one where the first school in the pair is included and the second school is removed, and another subsample where the second school is included and the first school is removed. When a school is removed from the sample, the weights of the remaining school are doubled to make up for the omitted school. In both subsamples, all students in the other sampling zones are included. With this process applied in each of the 75 sampling zones, the JRR procedure yields a total of 150 replicate subsamples, each one with its own set of replicate sampling weights to account for the successive removal of each school from the pair of schools in any given sampling zone.[3]

The process of creating replicate sampling weights for the replicate subsamples defines replicate factors k_{hj} as follows:

$$k_{hj} = \begin{cases} 2 \text{ for students in school } j \text{ of sampling zone } h \\ 0 \text{ for students in the other school of sampling zone } h \\ 1 \text{ for students in any other sampling zone} \end{cases} \tag{4.1}$$

These replicate factors are used to compute the 150 sets of replicate sampling weights as follows:

$$W_{hji} = k_{hj} \cdot W_{0i} \tag{4.2}$$

where W_{0i} is the overall sampling weight of student i and W_{hji} is the resulting replicate sampling weight of student i from sampling zone h when school j is included and the other school in the pair is removed.

Exhibit 4.2 illustrates how the replicate factors, necessary to produce the replicate sampling weights, are derived. Within each sampling zone, each school is assigned randomly an indicator u_{hj}, coded either 0 or 1, such that one school has a value of 0 and the other a value of 1. This indicator serves to identify which schools within each zone will be successively included or removed. When a school is removed from a zone, the replicate factor is set to zero and the sampling weights of all students in that school are set to zero; when a school is included, the replicate factor is set to two and the sampling weights of all students in that school are doubled. The sampling weights of students in all other sampling zones remain unchanged.

For example, sampling zone 1 yields two sets of replicate sampling weights. The first set has doubled sampling weights ($k_{11} = 2$) for the students in the first school ($u_{11} = 0$) of zone 1, zeroed sampling weights ($k_{12} = 0$) for the students in the second school ($u_{12} = 1$) of zone 1, and unchanged sampling weights ($k_{hj} = 1$) for all students in the other sampling zones. The second set of replicate sampling weights has zeroed sampling weights ($k_{11} = 0$) for the students in the first school ($u_{11} = 0$) of zone 1, doubled sampling weights ($k_{12} = 2$) for the students in the second school ($u_{12} = 1$) of zone 1, and unchanged sampling weights ($k_{hj} = 1$) for all students in the other sampling zones.

3 Prior to 2016, PIRLS used 75 subsamples and sets of replicate weights to calculate the JRR sampling variances. To provide more accurate estimates, starting in 2016 PIRLS uses 150 subsamples and sets of replicate weights to calculate the JRR sampling variances. Two subsamples are drawn from each sampling zone rather than one randomly selected subsample.

Exhibit 4.2: Construction of Replicate Factors Across Sampling Zones

Sample Zone	School Replicate Indicator (u_{hj})	Replicate Factors for Computing JRR Replicate Sampling Weights (k_{hj})											
		Zone 1		Zone 2		Zone 3		…	Zone h		…	Zone 75	
		(1)	(2)	(3)	(4)	(5)	(6)		(2h–1)	(2h)		(149)	(150)
1	0	2	0	1	1	1	1	…	1	1	…	1	1
	1	0	2										
2	0	1	1	2	0	1	1	…	1	1	…	1	1
	1			0	2								
3	0	1	1	1	1	2	0	…	1	1	…	1	1
	1					0	2						
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮	⋮
h	0	1	1	1	1	1	1	…	2	0	…	1	1
	1								0	2			
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮
75	0	1	1	1	1	1	1	…	1	1	…	2	0
	1											0	2

The process is repeated across all 75 possible sampling zones, generating 150 sets of replicate sampling weights. The replicate sampling weights are then used to estimate a statistic of interest 150 times. The variation across these 150 jackknife estimates determines the sampling variance.

Given a statistic t to be computed from a national sample, the formula used to estimate the sampling variance of that statistic, based on the PIRLS JRR algorithm, is given by the following equation:

$$Var_{jrr}(t_0) = \frac{1}{2} \sum_{h=1}^{75} \sum_{j=1}^{2} (t_{hj} - t_0)^2 \tag{4.3}$$

where the term t_0 denotes the statistic of interest estimated with the overall student sampling weights W_{0i} and the term t_{hj} denotes the same statistic computed using the set of replicate sampling weights W_{hji} obtained from sampling zone h (h=1,...,75), where the j^{th} school (1st or 2nd) in the zone is included and the other removed.

The sampling variance estimated with the PIRLS JRR method represents the variation arising from having sampled students using the multi-stage stratified cluster sample design. Its square root is the standard error for any statistic derived from variables other than plausible values. Examples of such statistics include the mean age of students, the mean scale score on the PIRLS *Students Like Reading* contextual scale, and the percentage of students that attended preprimary education three years or more.

Estimating Imputation Variance

For variables other than plausible values, standard errors were the result solely of sampling variation, and were computed using the JRR technique. However, the situation for plausible values was more complicated. As described in Chapter 2 of the *PIRLS 2016 Assessment Framework*, the PIRLS item pool was far too extensive to be administered in its entirety to any one student, and so a matrix-sampling assessment design was adopted whereby each student was given a single test booklet containing only a part of the entire assessment. The results for all of the booklets were then aggregated using item response theory (IRT) to provide results for the entire assessment. Multiple imputation was used to derive reliable estimates of student performance (plausible values) on the assessment as a whole, even though each student responded to just a subset of the assessment items. Because every student proficiency estimate incorporates a random element, PIRLS 2016 followed the customary procedure of generating five estimates for each student and using the variability among them as a measure of the imputation uncertainty, or error.

The general procedure for estimating the imputation variance when analyzing student achievement data follows the basic principle of performing any statistical analysis five times—once for each set of plausible values—and aggregating the five sets of results (Mislevy et al., 1992). Thus, for any given achievement-based statistic t, estimating that statistic from each plausible value yields five estimates t_m, $m = 1, \ldots, 5$, all of them computed using the overall student sampling weights W_{0i}. The final estimate of that statistic, t_0, is the average of these five estimates:

$$t_0 = \frac{1}{5}\sum_{m=1}^{5} t_m \tag{4.4}$$

The imputation variance of the statistic t_0 is simply the variance of the five results from the plausible values, computed as follows:

$$Var_{imp}(t_0) = \frac{6}{5}\sum_{m=1}^{5}\frac{(t_m - t_0)^2}{4} \tag{4.5}$$

where the factor $\frac{6}{5}$ is a correction factor required by the multiple imputation methodology. This imputation variance is then added to the sampling variance to produce the total variance estimate of the statistic t_0, as follows:

$$Var_{tot}(t_0) = Var_{jrr}(t_0) + Var_{imp}(t_0) \tag{4.6}$$

The sampling variance in this context is the average of the sampling variances from the five plausible values, as follows:

$$Var_{jrr}(t_0) = \frac{1}{5}\sum_{m=1}^{5} Var_{jrr}(t_m) \tag{4.7}$$

where

$$Var_{jrr}(t_m) = \frac{1}{2}\sum_{h=1}^{75}\sum_{j=1}^{2}(t_{mhj} - t_m)^2 \tag{4.8}$$

and t_{mhj} is the appropriate JRR estimate based on plausible value computed using the set of replicate sampling weights from sampling zone *h* where school *j* is included. The square root of the total variance is then the proper standard error for any statistic based on plausible values, such as the average PIRLS reading achievement for girls or the percentage of students who reached the PIRLS Advanced International Benchmark of reading achievement.

Appendices 4A and 4B provide details on the jackknife sampling variance, the imputation variance, the total variance, and the overall standard error for each country's mean proficiency estimates for PIRLS 2016 and ePIRLS 2016, respectively.

Estimating Standard Errors for International Averages

Some exhibits in the PIRLS 2016 reports include international averages and their standard errors, listed at the bottom of the exhibit. For example, Exhibit 1.5 of the *PIRLS 2016 International Results in Reading* report provides the international average for the percentages of girls and boys and their fourth grade reading achievement at the bottom of the exhibit. International averages are computed using the data from the participating countries included in the main table of an exhibit. Data from the benchmarking participants are not included in the estimation of international averages.

For any given statistic t_0, its international average is given by:

$$t_{int} = \frac{1}{N}\sum_{i=1}^{N} t_{0i} \tag{4.9}$$

where N is the number of countries contributing to the international average and t_{0i} is the estimate of our statistic of interest for the i^{th} country.

The variance of the international t_{int} average is given by:

$$Var(t_{int}) = \frac{1}{N^2}\sum_{i=1}^{N} Var_{tot}(t_{0i}) \tag{4.10}$$

where $Var_{tot}(t_{0i})$ is the total variance of our statistic of interest for the i^{th} country, as given in Equation (4.6) above. For statistics based on plausible values, the total variance includes the sampling variance and the imputation variance. For statistics not based on plausible values, such as percentages, the total variance is based entirely on the sampling variance, as shown in Equation (4.3) above. The standard error of the international average is the square root of the total variance.

References

Johnson, E.G., & Rust, K.F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics, 17*(2), 175–190.

Mislevy, R.J., Beaton, A., Kaplan, B.A., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29*(2), 133–161.

Wolter, K.M. (1985). *Introduction to variance estimation*. New York: Springer-Verlag.

Appendix 4A: Summary Statistics and Standard Errors for Proficiency in PIRLS Reading

Summary Statistics and Standard Errors for Proficiency in Overall Reading

Country	Sample Size	Overall Reading				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	6,341	544.360	5.802	0.614	6.416	2.533
Austria	4,360	540.796	5.297	0.370	5.667	2.381
Azerbaijan	5,994	472.277	16.798	0.650	17.447	4.177
Bahrain	5,480	445.999	4.549	0.854	5.403	2.325
Belgium (Flemish)	5,198	525.059	3.415	0.338	3.752	1.937
Belgium (French)	4,623	497.495	4.912	1.696	6.608	2.571
Bulgaria	4,281	551.539	17.215	0.794	18.009	4.244
Canada	18,245	543.098	3.182	0.165	3.348	1.830
Chile	4,294	493.872	5.408	0.812	6.220	2.494
Chinese Taipei	4,326	558.894	3.442	0.714	4.155	2.038
Czech Republic	5,537	543.348	4.245	0.244	4.488	2.119
Denmark	3,508	547.492	3.974	0.494	4.468	2.114
Egypt	6,957	330.471	28.680	3.194	31.874	5.646
England	5,095	558.682	3.461	0.117	3.578	1.891
Finland	4,896	566.007	3.090	0.311	3.400	1.844
France	4,767	511.244	4.510	0.225	4.736	2.176
Georgia	5,741	488.319	7.131	0.752	7.883	2.808
Germany	3,959	537.325	9.657	0.395	10.052	3.170
Hong Kong SAR	3,349	568.583	7.068	0.470	7.538	2.746
Hungary	4,623	554.160	8.178	0.114	8.292	2.880
Iran, Islamic Rep. of	8,766	427.899	13.543	2.065	15.608	3.951
Ireland	4,607	566.596	6.091	0.049	6.139	2.478
Israel	4,041	530.288	6.044	0.410	6.454	2.541
Italy	3,940	548.007	3.691	1.073	4.765	2.183
Kazakhstan	4,925	536.046	5.792	0.374	6.166	2.483
Kuwait	4,609	393.432	15.627	1.510	17.137	4.140
Latvia	4,157	557.751	2.724	0.158	2.882	1.698
Lithuania	4,317	548.278	5.969	0.929	6.898	2.626
Macao SAR	4,059	545.581	0.741	0.337	1.077	1.038
Malta	3,647	452.012	1.711	1.557	3.269	1.808

Summary Statistics and Standard Errors for Proficiency in Overall Reading (Continued)

Country	Sample Size	Overall Reading				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Morocco	10,942	357.823	13.813	1.694	15.506	3.938
Netherlands	4,206	544.884	2.670	0.132	2.803	1.674
New Zealand	5,646	522.531	4.517	0.224	4.741	2.177
Northern Ireland	3,693	564.621	4.201	0.701	4.902	2.214
Norway (5)	4,232	558.950	4.767	0.333	5.100	2.258
Oman	9,234	418.483	10.301	0.772	11.073	3.328
Poland	4,413	564.626	4.256	0.234	4.490	2.119
Portugal	4,642	527.797	4.944	0.254	5.198	2.280
Qatar	9,077	442.246	2.758	0.656	3.414	1.848
Russian Federation	4,577	580.772	4.397	0.448	4.846	2.201
Saudi Arabia	4,741	430.300	16.339	1.234	17.573	4.192
Singapore	6,488	576.178	9.769	0.171	9.940	3.153
Slovak Republic	5,451	534.791	8.555	1.163	9.719	3.117
Slovenia	4,499	542.466	3.640	0.267	3.907	1.977
South Africa	12,810	319.629	18.499	1.027	19.525	4.419
Spain	14,595	527.740	2.871	0.124	2.995	1.731
Sweden	4,525	555.160	5.550	0.195	5.745	2.397
Trinidad and Tobago	4,177	479.404	9.996	0.723	10.720	3.274
United Arab Emirates	16,471	450.104	9.815	0.641	10.456	3.234
United States	4,425	549.441	8.741	0.806	9.548	3.090
Benchmarking Participants						
Buenos Aires, Argentina	4,382	479.957	8.949	0.440	9.390	3.064
Ontario, Canada	4,270	543.582	10.050	0.225	10.276	3.206
Quebec, Canada	3,179	547.422	7.881	0.198	8.079	2.842
Denmark (3)	3,600	500.875	4.543	2.701	7.243	2.691
Norway (4)	4,354	516.874	3.283	0.593	3.876	1.969
Moscow City, Russian Fed.	4,289	612.084	4.170	0.476	4.646	2.155
Eng/Afr/Zulu - RSA (5)	5,282	406.012	34.276	1.391	35.667	5.972
Andalusia, Spain	4,169	524.584	4.113	0.232	4.345	2.084
Madrid, Spain	3,794	549.014	3.305	0.660	3.966	1.991
Abu Dhabi, UAE	4,188	414.308	20.632	1.567	22.199	4.712
Dubai, UAE	7,859	514.992	3.330	0.279	3.609	1.900

Summary Statistics and Standard Errors for Proficiency in Literary Experience

Country	Sample Size	Literary Experience				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	6,341	547.205	5.474	0.413	5.887	2.426
Austria	4,360	544.303	5.140	0.285	5.425	2.329
Azerbaijan	5,994	465.851	14.752	0.631	15.383	3.922
Bahrain	5,480	437.478	6.059	1.920	7.980	2.825
Belgium (Flemish)	5,198	523.594	3.637	0.111	3.748	1.936
Belgium (French)	4,623	503.821	4.786	0.139	4.925	2.219
Bulgaria	4,281	551.441	19.008	0.937	19.945	4.466
Canada	18,245	547.215	3.309	0.153	3.462	1.861
Chile	4,294	500.389	5.670	0.611	6.281	2.506
Chinese Taipei	4,326	548.387	3.265	0.907	4.172	2.042
Czech Republic	5,537	544.982	4.171	0.282	4.452	2.110
Denmark	3,508	551.281	3.972	0.709	4.681	2.163
Egypt	6,957	328.138	29.034	1.327	30.360	5.510
England	5,095	562.603	3.697	1.104	4.801	2.191
Finland	4,896	564.900	3.292	0.276	3.568	1.889
France	4,767	512.680	5.034	0.765	5.799	2.408
Georgia	5,741	489.905	6.860	0.071	6.931	2.633
Germany	3,959	542.338	9.951	0.875	10.826	3.290
Hong Kong SAR	3,349	562.473	7.404	1.505	8.909	2.985
Hungary	4,623	557.611	7.959	0.142	8.102	2.846
Iran, Islamic Rep. of	8,766	430.257	13.464	1.185	14.649	3.827
Ireland	4,607	571.308	6.148	1.111	7.259	2.694
Israel	4,041	532.226	6.399	0.291	6.690	2.587
Italy	3,940	548.737	4.037	0.406	4.442	2.108
Kazakhstan	4,925	527.236	5.512	0.810	6.322	2.514
Kuwait	4,609	387.778	16.543	1.820	18.363	4.285
Latvia	4,157	555.030	2.802	0.824	3.626	1.904
Lithuania	4,317	547.418	5.620	1.797	7.417	2.723
Macao SAR	4,059	535.999	0.892	1.954	2.846	1.687
Malta	3,647	451.899	2.075	1.890	3.965	1.991
Morocco	10,942	353.248	14.589	1.775	16.364	4.045
Netherlands	4,206	546.355	2.793	0.066	2.858	1.691

Summary Statistics and Standard Errors for Proficiency in Literary Experience (Continued)

Country	Sample Size	Literary Experience				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
New Zealand	5,646	525.278	4.731	0.650	5.381	2.320
Northern Ireland	3,693	570.464	4.563	1.599	6.162	2.482
Norway (5)	4,232	560.337	4.833	1.589	6.422	2.534
Oman	9,234	410.713	10.122	1.058	11.180	3.344
Poland	4,413	566.610	4.066	0.823	4.889	2.211
Portugal	4,642	527.774	5.208	0.906	6.115	2.473
Qatar	9,077	434.112	2.927	2.139	5.067	2.251
Russian Federation	4,577	579.129	4.237	0.471	4.708	2.170
Saudi Arabia	4,741	429.967	14.684	1.169	15.852	3.982
Singapore	6,488	574.559	9.542	1.193	10.735	3.276
Slovak Republic	5,451	538.758	8.747	0.407	9.154	3.026
Slovenia	4,499	541.192	3.653	2.152	5.805	2.409
South Africa	12,810	323.042	18.590	3.526	22.116	4.703
Spain	14,595	530.000	3.301	0.184	3.485	1.867
Sweden	4,525	555.953	5.513	0.231	5.744	2.397
Trinidad and Tobago	4,177	478.242	10.113	0.949	11.062	3.326
United Arab Emirates	16,471	439.953	10.606	0.638	11.244	3.353
United States	4,425	557.260	8.998	0.155	9.153	3.025
Benchmarking Participants						
Buenos Aires, Argentina	4,382	483.764	8.924	0.837	9.761	3.124
Ontario, Canada	4,270	548.600	10.074	0.398	10.473	3.236
Quebec, Canada	3,179	549.563	7.784	0.704	8.488	2.913
Denmark (3)	3,600	504.870	4.408	1.877	6.285	2.507
Norway (4)	4,354	520.498	3.437	0.783	4.220	2.054
Moscow City, Russian Fed.	4,289	613.262	4.095	0.915	5.010	2.238
Eng/Afr/Zulu - RSA (5)	5,282	401.912	37.153	2.567	39.720	6.302
Andalusia, Spain	4,169	525.589	4.068	0.274	4.342	2.084
Madrid, Spain	3,794	550.505	3.790	1.004	4.795	2.190
Abu Dhabi, UAE	4,188	405.509	22.119	0.938	23.058	4.802
Dubai, UAE	7,859	507.966	3.473	0.841	4.314	2.077

Summary Statistics and Standard Errors for Proficiency in Acquire and Use Information

Country	Sample Size	Acquire and Use Information				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	6,341	542.524	6.557	0.282	6.838	2.615
Austria	4,360	538.867	5.260	0.463	5.723	2.392
Azerbaijan	5,994	477.356	20.516	1.031	21.547	4.642
Bahrain	5,480	453.124	3.610	0.999	4.609	2.147
Belgium (Flemish)	5,198	525.831	3.297	0.431	3.728	1.931
Belgium (French)	4,623	490.108	5.189	0.429	5.619	2.370
Bulgaria	4,281	553.851	16.750	0.714	17.464	4.179
Canada	18,245	540.080	3.259	0.422	3.681	1.919
Chile	4,294	485.055	6.274	0.755	7.029	2.651
Chinese Taipei	4,326	569.214	3.530	1.157	4.687	2.165
Czech Republic	5,537	541.247	4.388	0.773	5.162	2.272
Denmark	3,508	543.284	4.565	1.737	6.302	2.510
Egypt	6,957	331.918	29.623	3.833	33.455	5.784
England	5,095	556.423	3.946	0.306	4.252	2.062
Finland	4,896	568.741	3.637	0.191	3.828	1.956
France	4,767	510.087	4.748	0.868	5.616	2.370
Georgia	5,741	486.383	8.775	0.990	9.765	3.125
Germany	3,959	532.921	9.859	1.197	11.056	3.325
Hong Kong SAR	3,349	576.386	7.236	0.358	7.594	2.756
Hungary	4,623	550.557	9.239	1.566	10.805	3.287
Iran, Islamic Rep. of	8,766	424.585	14.312	0.424	14.737	3.839
Ireland	4,607	564.727	6.584	0.454	7.038	2.653
Israel	4,041	528.681	6.168	0.044	6.212	2.492
Italy	3,940	548.960	3.737	1.203	4.940	2.223
Kazakhstan	4,925	543.594	6.649	1.361	8.010	2.830
Kuwait	4,609	398.428	18.028	0.489	18.517	4.303
Latvia	4,157	561.315	2.881	0.345	3.227	1.796
Lithuania	4,317	550.574	5.983	0.884	6.867	2.620
Macao SAR	4,059	555.505	0.779	0.892	1.671	1.293
Malta	3,647	451.399	1.438	2.558	3.996	1.999
Morocco	10,942	358.695	14.557	1.311	15.868	3.983
Netherlands	4,206	544.693	3.118	0.479	3.597	1.897

Summary Statistics and Standard Errors for Proficiency in Acquire and Use Information (Continued)

Country	Sample Size	Acquire and Use Information				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
New Zealand	5,646	520.376	4.869	0.851	5.720	2.392
Northern Ireland	3,693	560.782	4.273	1.094	5.368	2.317
Norway (5)	4,232	558.569	5.180	0.465	5.644	2.376
Oman	9,234	425.473	10.333	0.781	11.114	3.334
Poland	4,413	564.334	5.147	1.615	6.762	2.600
Portugal	4,642	528.388	5.065	0.335	5.400	2.324
Qatar	9,077	449.632	2.759	0.858	3.617	1.902
Russian Federation	4,577	584.419	4.669	0.398	5.068	2.251
Saudi Arabia	4,741	428.825	19.915	0.630	20.545	4.533
Singapore	6,488	578.591	10.208	0.724	10.932	3.306
Slovak Republic	5,451	531.052	9.387	0.399	9.786	3.128
Slovenia	4,499	544.294	3.864	0.459	4.323	2.079
South Africa	12,810	313.765	18.933	0.947	19.880	4.459
Spain	14,595	526.599	2.347	0.211	2.558	1.599
Sweden	4,525	554.850	6.242	0.550	6.792	2.606
Trinidad and Tobago	4,177	479.947	9.994	2.259	12.252	3.500
United Arab Emirates	16,471	459.769	9.255	0.958	10.214	3.196
United States	4,425	543.084	9.330	0.258	9.588	3.096
Benchmarking Participants						
Buenos Aires, Argentina	4,382	475.330	9.153	1.589	10.742	3.278
Ontario, Canada	4,270	539.458	11.135	0.659	11.794	3.434
Quebec, Canada	3,179	546.662	8.664	0.446	9.111	3.018
Denmark (3)	3,600	497.789	5.268	0.414	5.683	2.384
Norway (4)	4,354	513.681	3.523	1.179	4.701	2.168
Moscow City, Russian Fed.	4,289	613.081	4.901	1.550	6.450	2.540
Eng/Afr/Zulu - RSA (5)	5,282	407.024	35.042	0.725	35.766	5.981
Andalusia, Spain	4,169	523.891	4.030	0.831	4.860	2.205
Madrid, Spain	3,794	548.969	3.382	0.505	3.887	1.971
Abu Dhabi, UAE	4,188	422.034	19.937	4.633	24.570	4.957
Dubai, UAE	7,859	523.258	3.060	1.341	4.401	2.098

Summary Statistics and Standard Errors for Proficiency in Retrieving and Straightforward Inferencing

Country	Sample Size	Retrieving and Straightforward Inferencing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	6,341	540.737	5.620	0.979	6.599	2.569
Austria	4,360	550.244	6.319	1.601	7.920	2.814
Azerbaijan	5,994	477.348	16.328	1.155	17.483	4.181
Bahrain	5,480	444.491	4.017	0.462	4.478	2.116
Belgium (Flemish)	5,198	525.738	3.546	0.958	4.503	2.122
Belgium (French)	4,623	500.618	4.830	0.361	5.192	2.279
Bulgaria	4,281	550.435	16.093	0.103	16.196	4.024
Canada	18,245	541.431	3.109	0.106	3.215	1.793
Chile	4,294	495.873	5.478	0.606	6.084	2.466
Chinese Taipei	4,326	559.843	3.216	0.504	3.720	1.929
Czech Republic	5,537	551.194	4.669	1.051	5.720	2.392
Denmark	3,508	549.832	4.482	0.055	4.537	2.130
Egypt	6,957	329.082	29.228	2.052	31.280	5.593
England	5,095	555.703	3.678	0.305	3.983	1.996
Finland	4,896	572.066	3.660	0.249	3.908	1.977
France	4,767	520.580	4.772	0.477	5.249	2.291
Georgia	5,741	486.107	6.777	0.236	7.013	2.648
Germany	3,959	545.641	10.617	0.146	10.763	3.281
Hong Kong SAR	3,349	567.558	6.268	1.128	7.396	2.720
Hungary	4,623	551.523	8.892	1.821	10.714	3.273
Iran, Islamic Rep. of	8,766	429.490	14.304	1.796	16.100	4.012
Ireland	4,607	566.023	6.413	0.504	6.917	2.630
Israel	4,041	529.722	5.490	0.500	5.991	2.448
Italy	3,940	546.713	3.861	0.440	4.300	2.074
Kazakhstan	4,925	529.348	5.982	0.295	6.277	2.505
Kuwait	4,609	393.889	15.589	1.190	16.779	4.096
Latvia	4,157	554.060	2.938	0.786	3.724	1.930
Lithuania	4,317	549.379	5.557	0.945	6.503	2.550
Macao SAR	4,059	549.143	0.770	0.451	1.221	1.105
Malta	3,647	451.850	1.708	1.283	2.991	1.730
Morocco	10,942	363.775	13.018	2.071	15.089	3.884
Netherlands	4,206	546.451	3.070	0.929	3.999	2.000

Summary Statistics and Standard Errors for Proficiency in Retrieving and Straightforward Inferencing (Continued)

Country	Sample Size	Retrieving and Straightforward Inferencing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
New Zealand	5,646	521.368	4.408	1.014	5.421	2.328
Northern Ireland	3,693	561.558	4.251	0.304	4.554	2.134
Norway (5)	4,232	561.465	4.462	1.065	5.527	2.351
Oman	9,234	419.474	9.287	1.213	10.500	3.240
Poland	4,413	559.706	4.223	0.297	4.520	2.126
Portugal	4,642	527.791	4.700	0.243	4.943	2.223
Qatar	9,077	442.096	2.561	0.688	3.250	1.803
Russian Federation	4,577	581.389	4.941	0.210	5.152	2.270
Saudi Arabia	4,741	425.246	15.605	0.972	16.577	4.071
Singapore	6,488	573.013	9.375	0.274	9.649	3.106
Slovak Republic	5,451	537.543	8.881	0.533	9.414	3.068
Slovenia	4,499	546.631	4.628	0.435	5.063	2.250
South Africa	12,810	321.276	17.988	2.184	20.172	4.491
Spain	14,595	526.460	2.683	0.200	2.883	1.698
Sweden	4,525	560.141	6.157	0.997	7.154	2.675
Trinidad and Tobago	4,177	483.498	9.629	3.062	12.691	3.562
United Arab Emirates	16,471	448.078	9.133	1.151	10.283	3.207
United States	4,425	542.892	8.225	0.840	9.065	3.011
Benchmarking Participants						
Buenos Aires, Argentina	4,382	482.869	7.501	0.854	8.355	2.890
Ontario, Canada	4,270	538.853	9.940	0.749	10.689	3.269
Quebec, Canada	3,179	550.986	8.265	0.974	9.239	3.040
Denmark (3)	3,600	500.102	4.890	0.514	5.404	2.325
Norway (4)	4,354	521.395	3.411	0.617	4.028	2.007
Moscow City, Russian Fed.	4,289	611.229	4.378	1.182	5.559	2.358
Eng/Afr/Zulu - RSA (5)	5,282	407.415	34.318	3.110	37.428	6.118
Andalusia, Spain	4,169	522.016	3.549	0.081	3.630	1.905
Madrid, Spain	3,794	546.754	3.751	0.181	3.932	1.983
Abu Dhabi, UAE	4,188	412.982	19.278	1.499	20.776	4.558
Dubai, UAE	7,859	511.647	2.886	2.777	5.663	2.380

Summary Statistics and Standard Errors for Proficiency in Interpreting, Integrating, and Evaluating

Country	Sample Size	Interpreting, Integrating, and Evaluating				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Australia	6,341	549.233	5.487	0.116	5.602	2.367
Austria	4,360	534.439	5.227	1.174	6.401	2.530
Azerbaijan	5,994	464.716	18.646	0.251	18.896	4.347
Bahrain	5,480	445.507	4.668	2.651	7.319	2.705
Belgium (Flemish)	5,198	524.358	3.600	1.283	4.883	2.210
Belgium (French)	4,623	494.123	5.098	0.718	5.816	2.412
Bulgaria	4,281	552.315	18.098	0.383	18.481	4.299
Canada	18,245	545.111	3.146	0.204	3.350	1.830
Chile	4,294	491.144	6.137	2.357	8.494	2.914
Chinese Taipei	4,326	558.093	3.729	1.074	4.803	2.192
Czech Republic	5,537	537.904	4.207	0.648	4.855	2.203
Denmark	3,508	546.073	3.348	1.311	4.660	2.159
Egypt	6,957	339.914	26.558	6.019	32.577	5.708
England	5,095	561.489	3.453	0.156	3.609	1.900
Finland	4,896	562.473	3.246	0.054	3.300	1.817
France	4,767	501.030	5.169	0.376	5.545	2.355
Georgia	5,741	489.857	7.743	0.538	8.280	2.878
Germany	3,959	530.124	10.174	0.027	10.201	3.194
Hong Kong SAR	3,349	568.476	7.786	0.579	8.365	2.892
Hungary	4,623	556.735	8.693	0.367	9.060	3.010
Iran, Islamic Rep. of	8,766	424.686	13.694	2.953	16.647	4.080
Ireland	4,607	569.284	6.732	1.634	8.366	2.892
Israel	4,041	530.049	6.750	0.764	7.514	2.741
Italy	3,940	549.634	3.691	0.707	4.398	2.097
Kazakhstan	4,925	542.378	5.531	0.183	5.714	2.390
Kuwait	4,609	388.427	17.284	2.985	20.269	4.502
Latvia	4,157	561.800	2.574	0.239	2.813	1.677
Lithuania	4,317	547.824	5.935	1.025	6.960	2.638
Macao SAR	4,059	543.009	0.753	1.832	2.585	1.608
Malta	3,647	451.124	1.957	1.529	3.486	1.867
Morocco	10,942	336.140	17.391	2.660	20.052	4.478
Netherlands	4,206	544.387	2.649	0.359	3.008	1.734

Summary Statistics and Standard Errors for Proficiency in Interpreting, Integrating, and Evaluating (Continued)

Country	Sample Size	Interpreting, Integrating, and Evaluating				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
New Zealand	5,646	524.565	4.821	0.793	5.613	2.369
Northern Ireland	3,693	567.411	4.291	0.598	4.889	2.211
Norway (5)	4,232	558.138	4.888	0.879	5.767	2.401
Oman	9,234	414.785	11.307	1.473	12.781	3.575
Poland	4,413	569.549	4.132	1.472	5.604	2.367
Portugal	4,642	526.449	5.421	0.279	5.701	2.388
Qatar	9,077	440.920	2.627	0.918	3.544	1.883
Russian Federation	4,577	582.051	4.043	0.898	4.941	2.223
Saudi Arabia	4,741	438.563	16.027	1.174	17.200	4.147
Singapore	6,488	578.805	9.880	0.096	9.976	3.159
Slovak Republic	5,451	531.427	8.659	1.572	10.231	3.199
Slovenia	4,499	539.426	4.162	1.979	6.141	2.478
South Africa	12,810	308.245	20.834	6.739	27.572	5.251
Spain	14,595	529.100	2.729	0.251	2.979	1.726
Sweden	4,525	552.831	5.875	0.428	6.303	2.511
Trinidad and Tobago	4,177	472.284	11.842	1.377	13.219	3.636
United Arab Emirates	16,471	452.586	9.924	1.018	10.942	3.308
United States	4,425	554.964	8.973	0.651	9.624	3.102
Benchmarking Participants						
Buenos Aires, Argentina	4,382	472.816	11.927	1.640	13.567	3.683
Ontario, Canada	4,270	548.168	9.982	0.050	10.032	3.167
Quebec, Canada	3,179	545.009	8.457	0.526	8.983	2.997
Denmark (3)	3,600	503.905	5.126	1.026	6.152	2.480
Norway (4)	4,354	512.696	3.568	0.204	3.772	1.942
Moscow City, Russian Fed.	4,289	614.113	4.081	0.198	4.278	2.068
Eng/Afr/Zulu - RSA (5)	5,282	399.808	33.851	5.129	38.980	6.243
Andalusia, Spain	4,169	526.894	4.186	1.151	5.337	2.310
Madrid, Spain	3,794	549.928	3.759	0.129	3.888	1.972
Abu Dhabi, UAE	4,188	416.586	20.209	2.267	22.475	4.741
Dubai, UAE	7,859	518.784	3.165	0.469	3.634	1.906

Appendix 4B: Summary Statistics and Standard Errors for Proficiency in ePIRLS Online Informational Reading

Summary Statistics and Standard Errors for Proficiency in Overall ePIRLS Online Informational Reading

Country	Sample Size	Overall ePIRLS Online Informational Reading				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Canada	8,871	542.622	9.597	0.436	10.033	3.168
Chinese Taipei	4,299	545.648	3.651	0.279	3.930	1.982
Denmark	2,506	558.288	4.552	0.404	4.956	2.226
Georgia	5,557	476.903	10.138	0.613	10.752	3.279
Ireland	2,473	566.799	6.315	0.175	6.489	2.547
Israel	3,798	536.134	5.009	0.480	5.489	2.343
Italy	3,767	532.465	4.227	0.370	4.597	2.144
Norway (5)	3,610	567.537	4.316	0.597	4.913	2.217
Portugal	4,558	522.386	4.174	0.874	5.048	2.247
Singapore	6,320	588.129	9.024	0.147	9.171	3.028
Slovenia	4,303	525.010	3.153	0.649	3.802	1.950
Sweden	3,879	559.204	5.265	0.108	5.373	2.318
United Arab Emirates	15,566	468.330	4.365	0.475	4.840	2.200
United States	4,090	556.552	6.582	0.142	6.724	2.593
Benchmarking Participants						
Abu Dhabi, UAE	3,980	431.498	15.881	0.799	16.680	4.084
Dubai, UAE	7,471	527.726	2.015	0.505	2.520	1.588

Summary Statistics and Standard Errors for Proficiency in ePIRLS Retrieving and Straightforward Inferencing

Country	Sample Size	ePIRLS Retrieving and Straightforward Inferencing				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Canada	8,871	540.699	8.456	0.521	8.976	2.996
Chinese Taipei	4,299	548.349	4.059	0.560	4.619	2.149
Denmark	2,506	560.208	4.383	0.644	5.027	2.242
Georgia	5,557	484.799	10.799	0.365	11.164	3.341
Ireland	2,473	565.507	5.567	0.072	5.639	2.375
Israel	3,798	536.308	4.781	1.444	6.225	2.495
Italy	3,767	534.097	4.423	0.117	4.540	2.131
Norway (5)	3,610	567.395	4.441	0.470	4.911	2.216
Portugal	4,558	524.631	4.412	1.354	5.766	2.401
Singapore	6,320	594.394	10.090	0.597	10.687	3.269
Slovenia	4,303	525.401	3.086	0.194	3.280	1.811
Sweden	3,879	560.546	4.898	0.094	4.992	2.234
United Arab Emirates	15,566	470.775	4.320	0.286	4.606	2.146
United States	4,090	553.151	6.374	0.194	6.568	2.563
Benchmarking Participants						
Abu Dhabi, UAE	3,980	434.005	16.241	0.383	16.624	4.077
Dubai, UAE	7,471	527.813	2.229	0.809	3.038	1.743

Summary Statistics and Standard Errors for Proficiency in ePIRLS Interpreting, Integrating, and Evaluating

Country	Sample Size	ePIRLS Interpreting, Integrating, and Evaluating				
		Mean Proficiency	Jackknife Sampling Variance	Imputation Variance	Total Variance	Overall Standard Error
Canada	8,871	544.917	9.224	0.975	10.199	3.194
Chinese Taipei	4,299	543.919	3.423	0.052	3.475	1.864
Denmark	2,506	556.035	4.895	1.771	6.666	2.582
Georgia	5,557	465.986	11.288	2.513	13.800	3.715
Ireland	2,473	568.223	6.160	0.229	6.389	2.528
Israel	3,798	534.898	5.525	0.476	6.001	2.450
Italy	3,767	530.876	4.461	0.635	5.096	2.257
Norway (5)	3,610	567.564	4.686	0.452	5.138	2.267
Portugal	4,558	520.683	3.814	0.755	4.569	2.138
Singapore	6,320	584.729	8.428	1.119	9.547	3.090
Slovenia	4,303	523.398	3.737	0.142	3.879	1.970
Sweden	3,879	558.992	5.436	0.853	6.289	2.508
United Arab Emirates	15,566	465.079	4.230	0.430	4.660	2.159
United States	4,090	559.857	6.752	0.165	6.918	2.630
Benchmarking Participants						
Abu Dhabi, UAE	3,980	428.397	15.538	0.543	16.081	4.010
Dubai, UAE	7,471	527.349	1.724	0.883	2.607	1.615

CHAPTER 5

Sample Implementation in PIRLS 2016

Sylvie LaRoche
Pierre Foy

Overview

Rigorous sampling of schools and students was a key component of the PIRLS 2016 project. Implementing the sampling plan was the responsibility of the National Research Coordinator (NRC) in each participating country. NRCs were supported in this endeavor by the PIRLS 2016 sampling consultants, Statistics Canada, and the Sampling Unit of IEA Hamburg. Sampling consultants conducted the school sampling for most countries and trained NRCs using the Windows® Within-school Sampling Software (WinW3S) provided by IEA Hamburg to implement within-school sampling. As an essential part of their sampling activities, NRCs were responsible for providing detailed documentation describing their national sampling plans (sampling data, school sampling frames, and school sample selections). The documentation for each PIRLS participant was reviewed and completed by the sampling consultants, including detailed information on coverage and exclusion levels, stratification variables, sampling, participation rates, and variance estimates. The TIMSS & PIRLS International Study Center and the PIRLS 2016 Sampling Referee, Dr. Keith Rust of Westat, Inc., used this information to evaluate the quality of the samples.

This chapter provides a summary of the major characteristics of the national samples for PIRLS 2016, including PIRLS Literacy and ePIRLS. More detailed information on the sample design for each country, including details of population coverage and exclusions, stratification variables, and schools' sampling allocations, is provided in Appendix 5A Characteristics of National Samples.

Target Population

As described in Chapter 3 (Sample Design), the international target population for the PIRLS 2016 assessment is defined as the grade representing 4 years of formal schooling, counting from the first year of primary or elementary schooling.

For the PIRLS 2016 cycle, countries could participate in PIRLS Literacy—a less difficult reading assessment. PIRLS Literacy, which replaces prePIRLS from PIRLS 2011, was designed for countries where students found the PIRLS reading assessment too difficult. Countries considering PIRLS Literacy had the option of participating in PIRLS Literacy only or in both the PIRLS Literacy and PIRLS assessments. For countries who participated in both assessments, the student sample size was doubled and the PIRLS and PIRLS Literacy booklets were rotated within the sampled classes so that each student in the class was given either a PIRLS booklet or a PIRLS Literacy booklet.

The Islamic Republic of Iran and Morocco administered both PIRLS and PIRLS Literacy, while Egypt, Kuwait, and South Africa administered PIRLS Literacy only. Denmark administered PIRLS Literacy at the third grade and PIRLS at the fourth grade.

Exhibit 5.1 presents the grade identified as the target grade for sampling by each country and includes the number of years of formal schooling that the grades represent and the average age of students in the target grade at the time of testing.

For most countries, the target grade did indeed turn out to be the grade with 4 years of schooling—i.e., the fourth grade. However, in England, Malta, New Zealand, and Trinidad and Tobago, children begin primary school at an early age.[1] Therefore, these countries administered the PIRLS assessment in the fifth year of schooling. Norway chose to assess its fifth grade to obtain better comparisons with Sweden and Finland, while also assessing its fourth grade to measure trends to previous PIRLS assessments.

In addition to administering PIRLS Literacy at the fourth grade, South Africa administered PIRLS to assess students taught in English, Afrikaans, and Zulu at the fifth grade.

1 Given the cognitive demands of the assessment, PIRLS wants to avoid assessing very young students. Thus, PIRLS recommends assessing the next higher grade (i.e., fifth grade) if the average age at the time of testing would be less than 9.5 years.

Exhibit 5.1: National Grade Definition – PIRLS 2016

Country	Country's Name for Grade Tested	Years of Formal Schooling	Average Age at Time of Testing
Australia	Year 4	4	10.0
Austria	Grade 4	4	10.3
Azerbaijan	Grade 4	4	10.1
Bahrain	Grade 4	4	9.9
Belgium (Flemish)	Grade 4	4	10.1
Belgium (French)	Grade 4	4	10.0
Bulgaria	Grade 4	4	10.8
Canada	Grade 4	4	9.9
Chile	Grade 4	4	10.1
Chinese Taipei	Grade 4	4	10.1
Czech Republic	Grade 4	4	10.3
Denmark	Grade 4	4	10.8
Egypt	Grade 4	4	10.0
England	Year 5	5	10.3
Finland	Grade 4	4	10.8
France	Grade 4	4	9.8
Georgia	Grade 4	4	9.7
Germany	Grade 4	4	10.3
Hong Kong SAR	Primary 4	4	9.9
Hungary	Grade 4	4	10.6
Iran, Islamic Rep. of	Grade 4	4	10.2
Ireland	Fourth Class	4	10.5
Israel	Grade 4	4	10.0
Italy	Grade 4	4	9.7
Kazakhstan	Grade 4	4	10.3
Kuwait	Primary Grade 4	4	9.6
Latvia	Grade 4	4	10.9
Lithuania	Grade 4	4	10.8
Macao SAR	Primary 4	4	10.0
Malta	Year 5	5	9.7
Morocco	Grade 4	4	10.2
Netherlands	Grade 6	4	10.1
New Zealand	Year 5	4.5 - 5.5	10.1
Northern Ireland	Year 6	4	10.4
Norway (5)	Grade 5	5	10.8
Oman	Grade 4	4	9.7

Exhibit 5.1: National Grade Definition – PIRLS 2016 (Continued)

Country	Country's Name for Grade Tested	Years of Formal Schooling	Average Age at Time of Testing
Poland	Primary 4	4	10.7
Portugal	Grade 4	4	9.8
Qatar	Grade 5 for English curriculum schools; Grade 4 for other schools	4	10.0
Russian Federation	Grade 4	4	10.8
Saudi Arabia	Grade 4	4	9.9
Singapore	Grade 4	4	10.4
Slovak Republic	Grade 4	4	10.4
Slovenia	Grade 4	4	9.9
South Africa	Grade 4	4	10.6
Spain	Grade 4	4	9.9
Sweden	Grade 4	4	10.7
Trinidad and Tobago	Standard 3	5	10.2
United Arab Emirates	Grade 4	4	9.8
United States	Grade 4	4	10.1
Benchmarking Participants			
Buenos Aires, Argentina	Grade 4	4	10.0
Ontario, Canada	Grade 4	4	9.8
Quebec, Canada	Grade 4	4	10.1
Denmark (3)	Grade 3	3	9.8
Norway (4)	Grade 4	4	9.8
Moscow City, Russian Fed.	Grade 4	4	10.8
Eng/Afr/Zulu - RSA (5)	Grade 5	5	11.6
Andalusia, Spain	Grade 4	4	9.8
Madrid, Spain	Grade 4	4	9.9
Abu Dhabi, UAE	Grade 4	4	9.7
Dubai, UAE	Grade 4; Year 5 for schools following UK curriculum	4	9.9

National Coverage and Exclusions

Exhibits 5.2 summarizes population coverage and exclusions for the PIRLS 2016 and Exhibit 5.3 provides a similar summary for ePIRLS.

Coverage

National coverage of the PIRLS 2016 international target population was generally comprehensive, with some exceptions. These included Canada, which assessed students only from the provinces of Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland, Ontario, Quebec, and Saskatchewan, and Georgia, which assessed only students taught in Georgian and Azerbaijani. These participants chose a national target population that was less than the international target population. For these exceptions where coverage was below 100 percent, the results were footnoted in the PIRLS 2016 international reports.

The national coverage for PIRLS and ePIRLS was equivalent for every country but Canada. In Canada, only British Columbia, Newfoundland, Ontario, and Quebec took part in ePIRLS.

Exhibit 5.2: Coverage of Target Population – PIRLS 2016

	Country	International Target Population		Exclusions from National Target Population		
		Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
	Australia	100%		2.3%	2.4%	4.8%
2	Austria	100%		1.2%	4.4%	5.6%
	Azerbaijan	100%		2.1%	0.0%	2.1%
	Bahrain	100%		0.4%	2.3%	2.7%
	Belgium (Flemish)	100%		0.7%	0.9%	1.6%
2	Belgium (French)	100%		4.9%	1.1%	6.0%
	Bulgaria	100%		1.2%	3.1%	4.3%
1 2	Canada	97%	Students from the provinces of Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland, Ontario, Quebec, and Saskatchewan	2.8%	4.7%	7.5%
	Chile	100%		1.7%	2.3%	4.0%
	Chinese Taipei	100%		0.0%	0.9%	0.9%
	Czech Republic	100%		2.7%	0.7%	3.4%
2	Denmark	100%		1.9%	7.9%	9.8%
	Egypt	100%		1.2%	0.0%	1.2%
	England	100%		1.6%	2.1%	3.7%
	Finland	100%		1.3%	1.2%	2.4%
	France	100%		4.7%	0.6%	5.4%
1	Georgia	96%	Students taught in Georgian and Azerbaijani	0.8%	3.0%	3.8%
	Germany	100%		1.4%	2.8%	4.2%
2	Hong Kong SAR	100%		7.3%	2.8%	10.1%
	Hungary	100%		2.6%	1.9%	4.5%
	Iran, Islamic Rep. of	100%		3.9%	0.1%	4.1%
	Ireland	100%		2.3%	0.8%	3.1%
3	Israel	100%		21.0%	3.9%	24.9%
	Italy	100%		0.8%	4.1%	4.9%
	Kazakhstan	100%		4.1%	0.8%	4.9%
	Kuwait	100%		2.5%	1.4%	4.0%
2	Latvia	100%		4.3%	3.5%	7.9%
	Lithuania	100%		2.1%	2.1%	4.2%
	Macao SAR	100%		1.4%	2.2%	3.6%
2	Malta	100%		1.5%	6.4%	7.9%

1 National Target Population does not include all of the International Target Population.
2 National Defined Population covers 90% to 95% of National Target Population.
3 National Defined Population covers less than 90% of National Target Population (but at least 77%).

Exhibit 5.2: Coverage of Target Population – PIRLS 2016 (Continued)

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
Morocco	100%		1.7%	0.0%	1.7%
Netherlands	100%		2.4%	0.7%	3.1%
New Zealand	100%		1.3%	2.4%	3.7%
Northern Ireland	100%		2.6%	0.4%	3.0%
Norway (5)	100%		2.0%	3.3%	5.3%
Oman	100%		0.1%	0.5%	0.6%
Poland	100%		1.4%	2.5%	3.9%
[2] Portugal	100%		1.0%	6.5%	7.5%
Qatar	100%		2.0%	1.9%	3.9%
Russian Federation	100%		2.0%	2.1%	4.1%
Saudi Arabia	100%		1.9%	0.4%	2.3%
[3] Singapore	100%		10.6%	0.5%	11.1%
Slovak Republic	100%		3.1%	1.7%	4.8%
Slovenia	100%		1.5%	0.8%	2.4%
South Africa	100%		2.4%	0.2%	2.5%
Spain	100%		1.6%	3.2%	4.8%
Sweden	100%		1.3%	3.9%	5.2%
Trinidad and Tobago	100%		0.3%	1.0%	1.3%
United Arab Emirates	100%		2.0%	1.3%	3.3%
United States	100%		0.0%	4.8%	4.8%
Benchmarking Participants					
Buenos Aires, Argentina	100%		1.5%	1.2%	2.8%
Ontario, Canada	100%		2.3%	1.8%	4.1%
Quebec, Canada	100%		3.5%	1.6%	5.1%
[2] Denmark (3)	100%		1.9%	7.5%	9.3%
Norway (4)	100%		2.0%	3.0%	5.1%
Moscow City, Russian Fed.	100%		0.8%	2.6%	3.3%
Eng/Afr/Zulu - RSA (5)	100%		0.9%	0.2%	1.1%
Andalusia, Spain	100%		1.0%	3.2%	4.2%
[2] Madrid, Spain	100%		3.1%	3.4%	6.5%
Abu Dhabi, UAE	100%		2.2%	1.7%	3.9%
Dubai, UAE	100%		1.6%	1.5%	3.2%

Exhibit 5.3: Coverage of Target Population – ePIRLS 2016

Country	International Target Population		Exclusions from National Target Population		
	Coverage	Notes on Coverage	School-Level Exclusions	Within-Sample Exclusions	Overall Exclusions
1 2 Canada	74%	Students from the provinces of British Columbia, Newfoundland, Ontario, and Quebec	2.9%	3.6%	6.5%
Chinese Taipei	100%		0.0%	0.9%	0.9%
Denmark	100%		1.9%	8.0%	9.9%
1 Georgia	96%	Students taught in Georgian and Azerbaijani	0.8%	3.0%	3.8%
Ireland	100%		2.3%	1.4%	3.7%
3 Israel	100%		21.0%	3.9%	24.9%
Italy	100%		0.8%	4.1%	4.9%
Norway (5)	100%		2.0%	3.4%	5.3%
2 Portugal	100%		1.0%	6.5%	7.5%
3 Singapore	100%		10.6%	0.5%	11.1%
Slovenia	100%		1.5%	0.8%	2.4%
Sweden	100%		1.3%	3.9%	5.2%
United Arab Emirates	100%		2.0%	1.3%	3.3%
United States	100%		0.0%	4.9%	4.9%
Benchmarking Participants					
Abu Dhabi, UAE	100%		2.2%	1.7%	3.9%
Dubai, UAE	100%		1.6%	1.6%	3.2%

1 National Target Population does not include all of the International Target Population.
2 National Defined Population covers 90% to 95% of National Target Population.
3 National Defined Population covers less than 90% of National Target Population (but at least 77%).

School-Level and Student-Level Exclusions

Within the national target population, it was possible to exclude certain types of schools and students. For the most part, school-level exclusions consisted of schools for students with disabilities and very small or remote schools. Occasionally, schools were excluded for other reasons, as documented in Appendix 5A Characteristics of National Samples. Student-level, or within-school, exclusions generally consisted of students with disabilities or students who could not be assessed in the language of the test. For most PIRLS participants, the overall percentage of excluded students (combining school and within-school levels) was 5 percent or less after rounding. However, Austria, Belgium (French), Canada, Denmark, Hong Kong SAR, Latvia, Malta, and Portugal, as well as the benchmarking participants Denmark (3) and Madrid (Spain), had exclusions accounting for between 5 and 10 percent of the desired population. Israel and Singapore had exclusions exceeding 10 percent. Because the same students were sampled for ePIRLS in most countries, the ePIRLS overall exclusion rates were similar to those of PIRLS. Participants with an overall exclusion rate of more than 5 percent were annotated in the international reports.

Target Population Size

Exhibits 5.4 and 5.5 show the number of schools and students in each participant's target population[2] and sample for PIRLS and ePIRLS, respectively, as well as an estimate of the student population size based on the sample data. The target population figures were derived from the sampling frame used to select the PIRLS 2016 samples, and the sample figures were based on the number of sampled schools and students that participated in the assessments. The sample figures were computed using sampling weights (explained in more detail in Chapter 3). The student population size was based on the sampling frame and did not take into account the portion of the population excluded within sampled schools nor did it account for changes in the population between the date when the information in the sampling frame was collected and the date of the PIRLS 2016 data collection—usually a 2-year interval. Nevertheless, a comparison between the two estimates of population size can be seen as a validity check on the sampling procedure. In most cases, the population size estimated from the sample closely matched the population size from the sampling frame.

2 After school-level exclusions.

Exhibit 5.4: Population and Sample Sizes – PIRLS 2016

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated from Sample
Australia	6,530	275,099	286	6,341	287,196
Austria	3,020	81,005	150	4,360	81,450
Azerbaijan	3,709	122,286	170	5,994	128,877
Bahrain	183	17,769	182	5,480	17,493
Belgium (Flemish)	2,421	70,315	148	5,198	70,366
Belgium (French)	1,662	50,813	158	4,623	53,772
Bulgaria	1,752	62,074	153	4,281	60,411
Canada	9,377	344,011	926	18,245	342,617
Chile	6,012	228,629	154	4,294	230,972
Chinese Taipei	2,667	201,779	150	4,326	199,501
Czech Republic	3,440	102,460	157	5,537	99,938
Denmark	1,649	66,075	185	3,508	60,829
Egypt	16,401	1,610,893	160	6,957	1,543,299
England	14,946	597,669	170	5,095	588,313
Finland	2,237	58,254	151	4,896	55,611
France	31,577	776,184	163	4,767	787,106
Georgia	1,989	43,331	200	5,741	43,214
Germany	17,901	719,596	208	3,959	684,064
Hong Kong SAR	507	47,404	138	3,349	50,804
Hungary	2,796	91,826	149	4,623	90,647
Iran, Islamic Rep. of (Combined)	36,817	1,120,197	271	8,766	1,202,181
Literacy	36,817	1,120,197	271	4,381	1,202,181
PIRLS	36,817	1,120,197	271	4,385	1,202,181
Ireland	2,719	62,807	148	4,607	62,101
Israel	1,696	110,408	159	4,041	108,461
Italy	6,940	565,199	149	3,940	544,538
Kazakhstan	6,066	258,530	172	4,925	253,209
Kuwait	375	48,346	177	4,609	47,299
Latvia	649	18,515	150	4,157	18,478
Lithuania	827	25,969	195	4,317	25,062
Macao SAR	57	4,217	57	4,059	4,244
Malta	97	4,055	95	3,647	4,057

Exhibit 5.4: Population and Sample Sizes – PIRLS 2016 (Continued)

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated from Sample
Morocco (Combined)	19,216	649,390	360	10,942	664,737
Literacy	19,216	649,390	360	5,453	664,737
PIRLS	19,216	649,390	360	5,489	664,737
Netherlands	6,361	179,849	132	4,206	168,482
New Zealand	1,813	57,715	188	5,646	58,169
Northern Ireland	765	21,908	134	3,693	22,306
Norway (5)	1,991	59,159	150	4,232	58,583
Oman	662	54,975	306	9,234	52,512
Poland	11,473	368,742	148	4,413	333,001
Portugal	1,228	101,911	218	4,642	99,852
Qatar	208	19,690	216	9,077	19,791
Russian Federation	33,639	1,322,675	206	4,577	1,342,153
Saudi Arabia	11,708	438,538	202	4,741	433,654
Singapore	177	39,143	177	6,488	39,355
Slovak Republic	1,991	50,300	220	5,451	47,901
Slovenia	729	18,207	160	4,499	19,659
South Africa	16,896	944,645	293	12,810	983,873
Spain	12,730	473,955	629	14,595	472,876
Sweden	3,289	104,640	154	4,525	109,181
Trinidad and Tobago	511	18,956	151	4,177	18,333
United Arab Emirates	721	75,340	468	16,471	76,604
United States	69,235	3,989,251	158	4,425	3,752,434
Benchmarking Participants					
Buenos Aires, Argentina	876	38,886	150	4,382	41,023
Ontario, Canada	3,626	140,193	188	4,270	136,781
Quebec, Canada	1,726	75,398	127	3,179	74,775
Denmark (3)	1,649	66,075	186	3,600	62,709
Norway (4)	2,018	59,646	154	4,354	60,180
Moscow City, Russian Fed.	740	87,790	150	4,289	89,266
Eng/Afr/Zulu - RSA (5)	8,781	525,074	125	5,282	483,437
Andalusia, Spain	2,443	97,000	150	4,169	97,750
Madrid, Spain	1,293	66,613	168	3,794	65,346
Abu Dhabi, UAE	278	26,871	151	4,188	27,825
Dubai, UAE	161	20,920	174	7,859	21,867

Exhibit 5.5: Population and Sample Sizes – ePIRLS 2016

Country	Population		Sample		
	Schools	Students	Schools	Students	Student Population Size Estimated from Sample
Canada	9,902	262,540	474	8,871	264,737
Chinese Taipei	2,667	201,779	150	4,299	199,501
Denmark	1,649	66,075	142	2,506	60,103
Georgia	1,989	43,331	199	5,557	43,210
Ireland	2,719	62,807	147	2,473	62,393
Israel	1,696	110,408	157	3,798	108,348
Italy	6,940	565,199	148	3,767	544,871
Norway (5)	1,991	59,159	142	3,610	58,862
Portugal	1,228	101,911	218	4,558	99,852
Singapore	177	39,143	177	6,320	39,355
Slovenia	729	18,207	159	4,303	19,668
Sweden	3,289	104,640	144	3,879	109,160
United Arab Emirates	721	75,340	465	15,566	76,653
United States	69,235	3,989,251	153	4,090	3,765,069
Benchmarking Participants					
Abu Dhabi, UAE	278	26,871	150	3,980	27,869
Dubai, UAE	161	20,920	174	7,741	21,895

Meeting PIRLS 2016 Standards for Sampling Participation

PIRLS 2016 participants understood that the goal for sampling participation was 100 percent for all sampled schools, classrooms, and students. Guidelines for reporting achievement data for participants that secure less than full participation were modeled after IEA's previous PIRLS assessment cycles. As summarized below in Exhibit 5.6, countries were assigned to one of three categories on the basis of their sampling participation. Countries in Category 1 were considered to have met all PIRLS 2016 sampling requirements and to have acceptable participation rates. Countries in Category 2 met the participation requirements only after including replacement schools. Countries that failed to meet the participation requirements even with the use of replacement schools were assigned to Category 3. One of the main goals for quality data in PIRLS 2016 was to have as many countries as possible achieve Category 1 status.

Exhibit 5.6: Categories of Sampling Participation

Category 1	Acceptable sampling participation rate **without** the use of replacement schools. In order to be placed in this category, a country had to have: ▪ An **unweighted** school response rate **without** replacement of at least 85% (after rounding to nearest whole percent) AND an **unweighted** student response rate (after rounding) of at least 85% OR ▪ A **weighted** school response rate **without** replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR ▪ The product of the (unrounded) **weighted** school response rate **without** replacement and the (unrounded) **weighted** student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category would appear in the tables and figures in international reports without annotation, and will be ordered by achievement as appropriate.
Category 2	Acceptable sampling participation rate **only when replacement schools are included**. A country would be placed in this category 2 if: ▪ It failed to meet the requirements for Category 1 but had a weighted school response rate **without** replacement of at least 50% (after rounding to the nearest percent) AND HAD EITHER ▪ A **weighted** school response rate **with** replacement of at least 85% (after rounding to nearest whole percent) AND a **weighted** student response rate (after rounding) of at least 85% OR ▪ The product of the (unrounded) **weighted** school response rate **with** replacement and the (unrounded) **weighted** student response rate of at least 75% (after rounding to the nearest whole percent). Countries in this category would be annotated with † in the tables and figures in international reports, and ordered by achievement as appropriate.
Category 3	Unacceptable sampling response rate even when replacement schools are included. Countries that could provide documentation to show that they complied with PIRLS sampling procedures and requirements but did not meet the requirements for Category 1 or Category 2 would be placed in Category 3. Countries in this category would be annotated with ‡ if they nearly met the requirements for Category 2. Countries would be annotated with ≡ if they failed to meet the participation requirements but had a school participation rate of at least 50% before the use of replacement schools. At last, if none of these conditions are met, countries would appear in a separate section of the achievement tables, below the other countries, in international reports. These countries would be presented in alphabetical order.

Exhibits 5.7 and 5.8 present the weighted school, classroom, student, and overall participation rates in the PIRLS and ePIRLS assessments, and Exhibits 5.9 and 5.10 present the unweighted participation rates. Almost all PIRLS participants had excellent participation rates and were classified as Category 1. Hong Kong SAR, the Netherlands, and the United States achieved the minimum acceptable participation rate only after including replacement schools, and therefore their results were annotated with the symbol † in the achievement exhibits of the PIRLS international results report (Category 2). Despite efforts to secure full participation, the benchmarking

participant Quebec, Canada, did not meet the required sampling participation rate even with the use of replacement schools and was annotated with the symbol ≡ in the achievement exhibits of the report (Category 3).

Similarly, nearly all ePIRLS participants had very good participation rates and were classified as Category 1. The United States achieved the minimum acceptable participation rate only after including replacement schools and were annotated with the symbol † in the achievement exhibits of the ePIRLS report (Category 2). In spite of efforts to achieve full participation, Denmark did not meet the required sampling participation rate in ePIRLS even with the replacement schools and their achievement results were annotated with the symbol ≡ in the report (Category 3).

Exhibit 5.7: Participation Rates (Weighted) – PIRLS 2016

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Australia	97%	100%	100%	95%	92%	94%
Austria	100%	100%	100%	98%	98%	98%
Azerbaijan	100%	100%	100%	96%	96%	96%
Bahrain	99%	99%	100%	98%	98%	98%
Belgium (Flemish)	79%	94%	100%	98%	77%	92%
Belgium (French)	96%	100%	100%	97%	93%	97%
Bulgaria	100%	100%	100%	95%	95%	95%
Canada	81%	90%	100%	96%	77%	86%
Chile	92%	100%	100%	96%	88%	96%
Chinese Taipei	100%	100%	100%	98%	98%	98%
Czech Republic	100%	100%	100%	95%	95%	95%
Denmark	87%	96%	100%	94%	82%	90%
Egypt	100%	100%	100%	97%	97%	97%
England	99%	100%	100%	96%	95%	96%
Finland	98%	99%	100%	96%	95%	96%
France	99%	100%	100%	96%	95%	96%
Georgia	98%	99%	100%	97%	95%	96%
Germany	97%	100%	100%	96%	93%	95%
† Hong Kong SAR	74%	91%	100%	87%	64%	79%
Hungary	98%	100%	100%	97%	95%	97%
Iran, Islamic Rep. of (Combined)	100%	100%	100%	99%	99%	99%
Literacy	100%	100%	100%	99%	99%	99%
PIRLS	100%	100%	100%	99%	99%	99%
Ireland	100%	100%	100%	96%	96%	96%
Israel	98%	99%	100%	95%	93%	94%
Italy	89%	99%	100%	96%	85%	95%
Kazakhstan	100%	100%	100%	99%	99%	99%
Kuwait	98%	98%	100%	93%	91%	91%
Latvia	95%	97%	100%	94%	89%	91%
Lithuania	100%	100%	100%	95%	95%	95%
Macao SAR	100%	100%	100%	98%	98%	98%
Malta	100%	100%	100%	96%	96%	96%

PIRLS guidelines for sampling participation: The minimum acceptable participation rates were 85 percent of both schools and students, or a combined rate (the product of school and student participation) of 75 percent. Participants not meeting these guidelines were annotated as follows:

† Met guidelines for sample participation rates only after replacement schools were included.

‡ Nearly satisfied guidelines for sample participation rates after replacement schools were included.

≡ Did not satisfy guidelines for sample participation rates.

Exhibit 5.7: Participation Rates (Weighted) – PIRLS 2016 (Continued)

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Morocco (Combined)	100%	100%	100%	99%	99%	99%
Literacy	100%	100%	100%	99%	99%	99%
PIRLS	100%	100%	100%	99%	99%	99%
† Netherlands	69%	90%	100%	96%	66%	86%
New Zealand	85%	97%	100%	96%	81%	92%
Northern Ireland	84%	88%	100%	96%	81%	84%
Norway (5)	95%	99%	100%	96%	91%	95%
Oman	99%	100%	100%	99%	98%	98%
Poland	95%	99%	100%	91%	86%	90%
Portugal	97%	99%	100%	94%	91%	93%
Qatar	100%	100%	100%	97%	97%	97%
Russian Federation	100%	100%	100%	98%	98%	98%
Saudi Arabia	92%	100%	100%	96%	88%	96%
Singapore	100%	100%	100%	97%	97%	97%
Slovak Republic	94%	100%	100%	97%	92%	97%
Slovenia	94%	94%	100%	96%	90%	90%
South Africa	92%	97%	100%	96%	88%	94%
Spain	99%	100%	100%	97%	95%	97%
Sweden	99%	100%	100%	95%	94%	95%
Trinidad and Tobago	100%	100%	100%	96%	96%	96%
United Arab Emirates	98%	99%	100%	96%	95%	95%
† United States	75%	92%	100%	94%	71%	86%
Benchmarking Participants						
Buenos Aires, Argentina	88%	100%	100%	92%	81%	92%
Ontario, Canada	96%	97%	100%	96%	92%	93%
≡ Quebec, Canada	39%	67%	99%	96%	37%	64%
Denmark (3)	88%	97%	100%	95%	83%	92%
Norway (4)	95%	99%	100%	96%	91%	95%
Moscow City, Russian Fed.	100%	100%	100%	97%	97%	97%
Eng/Afr/Zulu - RSA (5)	84%	89%	100%	96%	81%	86%
Andalusia, Spain	99%	100%	100%	96%	96%	96%
Madrid, Spain	100%	100%	100%	97%	97%	97%
Abu Dhabi, UAE	100%	100%	100%	96%	96%	96%
Dubai, UAE	99%	99%	100%	96%	95%	95%

Exhibit 5.8: **Participation Rates (Weighted) – ePIRLS 2016**

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Canada	79%	85%	100%	93%	74%	79%
Chinese Taipei	100%	100%	100%	98%	98%	98%
≡ Denmark	67%	74%	100%	87%	58%	64%
Georgia	97%	99%	100%	95%	92%	94%
Ireland	99%	99%	100%	91%	91%	91%
Israel	97%	98%	100%	91%	88%	89%
Italy	89%	99%	100%	92%	82%	91%
Norway (5)	91%	93%	99%	88%	79%	81%
Portugal	97%	99%	100%	92%	90%	91%
Singapore	100%	100%	100%	95%	95%	95%
Slovenia	94%	94%	99%	93%	86%	86%
Sweden	93%	93%	99%	90%	83%	83%
United Arab Emirates	98%	98%	100%	92%	90%	90%
† United States	74%	89%	100%	90%	67%	80%
Benchmarking Participants						
Abu Dhabi, UAE	99%	99%	100%	92%	91%	91%
Dubai, UAE	99%	99%	99%	92%	91%	91%

PIRLS guidelines for sampling participation: The minimum acceptable participation rates were 85 percent of both schools and students, or a combined rate (the product of school and student participation) of 75 percent. Participants not meeting these guidelines were annotated as follows:

† Met guidelines for sample participation rates only after replacement schools were included.

‡ Nearly satisfied guidelines for sample participation rates after replacement schools were included.

≡ Did not satisfy guidelines for sample participation rates.

Exhibit 5.9: **Participation Rates (Unweighted) – PIRLS 2016**

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Australia	98%	100%	97%	94%	89%	91%
Austria	100%	100%	100%	98%	98%	98%
Azerbaijan	100%	100%	100%	96%	96%	96%
Bahrain	99%	99%	100%	98%	98%	98%
Belgium (Flemish)	79%	94%	100%	98%	77%	92%
Belgium (French)	96%	100%	100%	97%	93%	97%
Bulgaria	100%	100%	100%	95%	95%	95%
Canada	87%	93%	100%	96%	83%	89%
Chile	90%	100%	100%	96%	86%	96%
Chinese Taipei	100%	100%	100%	98%	98%	98%
Czech Republic	100%	100%	100%	95%	95%	95%
Denmark	89%	97%	100%	94%	83%	91%
Egypt	100%	100%	100%	97%	97%	97%
England	99%	100%	100%	96%	95%	96%
Finland	98%	99%	100%	96%	94%	96%
France	99%	100%	100%	96%	95%	96%
Georgia	99%	100%	100%	97%	95%	96%
Germany	98%	100%	100%	96%	94%	96%
Hong Kong SAR	75%	91%	100%	86%	65%	78%
Hungary	98%	100%	100%	97%	95%	97%
Iran, Islamic Rep. of (Combined)	100%	100%	100%	99%	99%	99%
Literacy	100%	100%	100%	99%	99%	99%
PIRLS	100%	100%	100%	99%	99%	99%
Ireland	100%	100%	100%	96%	96%	96%
Israel	98%	99%	100%	95%	93%	95%
Italy	89%	99%	100%	96%	85%	95%
Kazakhstan	99%	100%	100%	99%	98%	99%
Kuwait	98%	98%	100%	92%	90%	90%
Latvia	94%	97%	100%	93%	87%	90%
Lithuania	100%	100%	100%	95%	95%	95%
Macao SAR	100%	100%	100%	98%	98%	98%
Malta	100%	100%	100%	96%	96%	96%
Morocco (Combined)	100%	100%	100%	98%	98%	98%
Literacy	100%	100%	100%	98%	98%	98%
PIRLS	100%	100%	100%	98%	98%	98%
Netherlands	68%	89%	100%	96%	65%	85%

Exhibit 5.9: Participation Rates (Unweighted) – PIRLS 2016 (Continued)

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
New Zealand	84%	95%	100%	95%	80%	90%
Northern Ireland	85%	88%	100%	95%	81%	84%
Norway (5)	95%	99%	100%	96%	92%	95%
Oman	99%	100%	100%	98%	98%	98%
Poland	95%	99%	100%	90%	85%	89%
Portugal	95%	99%	100%	94%	89%	92%
Qatar	100%	100%	100%	97%	97%	97%
Russian Federation	100%	100%	100%	98%	98%	98%
Saudi Arabia	92%	100%	100%	95%	87%	95%
Singapore	100%	100%	100%	97%	97%	97%
Slovak Republic	95%	100%	100%	97%	92%	97%
Slovenia	94%	94%	100%	96%	91%	91%
South Africa	93%	97%	100%	96%	90%	93%
Spain	99%	100%	100%	97%	96%	97%
Sweden	99%	100%	100%	95%	94%	95%
Trinidad and Tobago	100%	100%	100%	96%	96%	96%
United Arab Emirates	98%	99%	100%	97%	95%	95%
United States	76%	92%	100%	94%	71%	86%
Benchmarking Participants						
Buenos Aires, Argentina	87%	100%	100%	92%	80%	92%
Ontario, Canada	95%	96%	100%	96%	91%	92%
Quebec, Canada	51%	73%	99%	96%	48%	69%
Denmark (3)	89%	97%	100%	95%	84%	92%
Norway (4)	95%	99%	100%	96%	91%	95%
Moscow City, Russian Fed.	100%	100%	100%	97%	97%	97%
Eng/Afr/Zulu - RSA (5)	84%	90%	100%	96%	81%	87%
Andalusia, Spain	99%	100%	100%	97%	95%	97%
Madrid, Spain	100%	100%	100%	97%	97%	97%
Abu Dhabi, UAE	100%	100%	100%	96%	96%	96%
Dubai, UAE	99%	99%	100%	96%	96%	96%

Exhibit 5.10: Participation Rates (Unweighted) – ePIRLS 2016

Country	School Participation		Class Participation	Student Participation	Overall Participation	
	Before Replacement	After Replacement			Before Replacement	After Replacement
Canada	93%	94%	100%	91%	85%	86%
Chinese Taipei	100%	100%	100%	98%	98%	98%
Denmark	69%	74%	100%	87%	60%	65%
Georgia	98%	99%	100%	94%	93%	93%
Ireland	99%	99%	100%	91%	91%	91%
Israel	97%	98%	100%	91%	88%	89%
Italy	89%	99%	100%	92%	81%	90%
Norway (5)	91%	93%	99%	88%	79%	81%
Portugal	95%	99%	100%	92%	88%	91%
Singapore	100%	100%	100%	94%	94%	94%
Slovenia	94%	94%	99%	93%	86%	86%
Sweden	94%	94%	98%	90%	82%	82%
United Arab Emirates	98%	98%	99%	92%	90%	90%
United States	74%	89%	100%	90%	67%	80%
Benchmarking Participants						
Abu Dhabi, UAE	99%	99%	99%	92%	91%	91%
Dubai, UAE	99%	99%	99%	92%	91%	91%

Exhibits 5.11 and 5.12 show the achieved sample sizes in terms of schools for each of the participants in the PIRLS and ePIRLS assessments, respectively, and Exhibits 5.13 and 5.14 show the achieved sample sizes on these assessments in terms of students.

Exhibit 5.11: School Sample Sizes – PIRLS 2016

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Australia	286	286	281	5	286
Austria	152	150	150	0	150
Azerbaijan	170	170	170	0	170
Bahrain	184	183	182	0	182
Belgium (Flemish)	160	157	124	24	148
Belgium (French)	158	158	152	6	158
Bulgaria	154	153	153	0	153
Canada	1,020	998	872	54	926
Chile	154	154	139	15	154
Chinese Taipei	150	150	150	0	150
Czech Republic	157	157	157	0	157
Denmark	198	191	170	15	185
Egypt	160	160	160	0	160
England	171	170	168	2	170
Finland	159	152	149	2	151
France	166	163	161	2	163
Georgia	201	201	198	2	200
Germany	210	209	204	4	208
Hong Kong SAR	152	151	114	24	138
Hungary	154	149	146	3	149
Iran, Islamic Rep. of	274	271	271	0	271
Ireland	150	148	148	0	148
Israel	160	160	157	2	159
Italy	150	150	134	15	149
Kazakhstan	174	172	171	1	172
Kuwait	187	181	177	0	177
Latvia	156	154	145	5	150
Lithuania	196	195	195	0	195
Macao SAR	57	57	57	0	57
Malta	97	95	95	0	95
Morocco	361	360	360	0	360
Netherlands	150	148	101	31	132
New Zealand	198	198	167	21	188
Northern Ireland	154	153	130	4	134

Exhibit 5.11: School Sample Sizes – PIRLS 2016 (Continued)

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Norway (5)	153	152	145	5	150
Oman	308	307	305	1	306
Poland	150	149	141	7	148
Portugal	222	221	211	7	218
Qatar	218	216	216	0	216
Russian Federation	206	206	206	0	206
Saudi Arabia	208	202	185	17	202
Singapore	177	177	177	0	177
Slovak Republic	221	220	208	12	220
Slovenia	172	170	160	0	160
South Africa	304	302	282	11	293
Spain	630	629	625	4	629
Sweden	158	154	153	1	154
Trinidad and Tobago	152	151	151	0	151
United Arab Emirates	482	475	467	1	468
United States	176	172	131	27	158
Benchmarking Participants					
Buenos Aires, Argentina	150	150	131	19	150
Ontario, Canada	198	196	186	2	188
Quebec, Canada	176	174	89	38	127
Denmark (3)	198	191	170	16	186
Norway (4)	155	155	147	7	154
Moscow City, Russian Fed.	150	150	150	0	150
Eng/Afr/Zulu - RSA (5)	152	139	117	8	125
Andalusia, Spain	150	150	148	2	150
Madrid, Spain	168	168	168	0	168
Abu Dhabi, UAE	153	151	151	0	151
Dubai, UAE	178	175	174	0	174

Exhibit 5.12: School Sample Sizes – ePIRLS 2016

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample that Participated	Number of Replacement Schools that Participated	Total Number of Schools that Participated
Canada	507	503	467	7	474
Chinese Taipei	150	150	150	0	150
Denmark	198	191	132	10	142
Georgia	201	201	197	2	199
Ireland	150	148	147	0	147
Israel	160	160	155	2	157
Italy	150	150	133	15	148
Norway (5)	153	152	138	4	142
Portugal	222	221	211	7	218
Singapore	177	177	177	0	177
Slovenia	172	170	159	0	159
Sweden	158	154	144	0	144
United Arab Emirates	482	475	464	1	465
United States	176	172	128	25	153
Benchmarking Participants					
Abu Dhabi, UAE	153	151	150	0	150
Dubai, UAE	178	175	174	0	174

Exhibit 5.13: Student Sample Sizes – PIRLS 2016

Country	Within-school Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Eligible Students	Number of Students Absent	Number of Students Assessed
Australia	95%	7,064	168	155	6,741	400	6,341
Austria	98%	4,709	20	222	4,467	107	4,360
Azerbaijan	96%	6,361	113	0	6,248	254	5,994
Bahrain	98%	5,771	56	148	5,567	87	5,480
Belgium (Flemish)	98%	5,378	39	28	5,311	113	5,198
Belgium (French)	97%	4,841	8	64	4,769	146	4,623
Bulgaria	95%	4,677	75	108	4,494	213	4,281
Canada	96%	20,072	265	736	19,071	826	18,245
Chile	96%	4,648	73	85	4,490	196	4,294
Chinese Taipei	98%	4,471	39	38	4,394	68	4,326
Czech Republic	95%	5,939	78	35	5,826	289	5,537
Denmark	94%	4,091	68	278	3,745	237	3,508
Egypt	97%	7,321	150	0	7,171	214	6,957
England	96%	5,568	149	113	5,306	211	5,095
Finland	96%	5,178	52	42	5,084	188	4,896
France	96%	5,050	56	33	4,961	194	4,767
Georgia	97%	6,123	59	131	5,933	192	5,741
Germany	96%	4,279	58	102	4,119	160	3,959
Hong Kong SAR	87%	4,024	21	96	3,907	558	3,349
Hungary	97%	4,852	21	57	4,774	151	4,623
Iran, Islamic Rep. of (Combined)	99%	8,999	106	10	8,883	117	8,766
Literacy	99%	4,498	53	4	4,441	60	4,381
PIRLS	99%	4,501	53	6	4,442	57	4,385
Ireland	96%	4,881	30	44	4,807	200	4,607
Israel	95%	4,368	13	107	4,248	207	4,041
Italy	96%	4,309	22	166	4,121	181	3,940
Kazakhstan	99%	5,035	51	0	4,984	59	4,925
Kuwait	93%	5,082	66	14	5,002	393	4,609
Latvia	94%	4,636	21	134	4,481	324	4,157
Lithuania	95%	4,670	35	79	4,556	239	4,317
Macao SAR	98%	4,254	10	93	4,151	92	4,059
Malta	96%	4,022	6	223	3,793	146	3,647

Students attending a sampled class at the time the sample was chosen but leaving the class before the assessment was administered were classified as "withdrawn."
Students with a disability or language barrier that prevented them from participating in the assessment were classified as "excluded."
Students not present when the assessment was administered, and not subsequently assessed in a make-up session, were classified as "absent."

Exhibit 5.13: Student Sample Sizes – PIRLS 2016 (Continued)

Country	Within-school Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Eligible Students	Number of Students Absent	Number of Students Assessed
Morocco (Combined)	99%	11,370	194	0	11,176	234	10,942
Literacy	99%	5,680	94	0	5,586	133	5,453
PIRLS	99%	5,690	100	0	5,590	101	5,489
Netherlands	96%	4,446	42	15	4,389	183	4,206
New Zealand	96%	6,128	77	119	5,932	286	5,646
Northern Ireland	96%	3,920	27	20	3,873	180	3,693
Norway (5)	96%	4,595	49	142	4,404	172	4,232
Oman	99%	9,619	146	67	9,406	172	9,234
Poland	91%	5,069	43	125	4,901	488	4,413
Portugal	94%	5,305	58	293	4,954	312	4,642
Qatar	97%	9,730	182	205	9,343	266	9,077
Russian Federation	98%	4,740	4	63	4,673	96	4,577
Saudi Arabia	96%	5,044	37	23	4,984	243	4,741
Singapore	97%	6,719	29	0	6,690	202	6,488
Slovak Republic	97%	5,869	207	41	5,621	170	5,451
Slovenia	96%	4,721	10	35	4,676	177	4,499
South Africa	96%	13,669	348	26	13,295	485	12,810
Spain	97%	15,634	55	520	15,059	464	14,595
Sweden	95%	4,988	38	189	4,761	236	4,525
Trinidad and Tobago	96%	4,506	108	50	4,348	171	4,177
United Arab Emirates	96%	17,381	89	232	17,060	589	16,471
United States	94%	5,056	159	175	4,722	297	4,425
Benchmarking Participants							
Buenos Aires, Argentina	92%	4,843	46	43	4,754	372	4,382
Ontario, Canada	96%	4,572	50	71	4,451	181	4,270
Quebec, Canada	96%	3,396	17	59	3,320	141	3,179
Denmark (3)	95%	4,120	60	261	3,799	199	3,600
Norway (4)	96%	4,725	46	138	4,541	187	4,354
Moscow City, Russian Fed.	97%	4,494	14	49	4,431	142	4,289
Eng/Afr/Zulu - RSA (5)	96%	5,692	197	16	5,479	197	5,282
Andalusia, Spain	96%	4,470	22	132	4,316	147	4,169
Madrid, Spain	97%	4,050	16	127	3,907	113	3,794
Abu Dhabi, UAE	96%	4,408	20	27	4,361	173	4,188
Dubai, UAE	96%	8,356	50	148	8,158	299	7,859

Exhibit 5.14: Student Sample Sizes – ePIRLS 2016

Country	Within-school Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Eligible Students	Number of Students Absent	Number of Students Assessed
Canada	93%	10,178	83	391	9,704	833	8,871
Chinese Taipei	98%	4,471	39	38	4,394	95	4,299
Denmark	87%	3,139	48	219	2,872	366	2,506
Georgia	95%	6,072	58	128	5,886	329	5,557
Ireland	91%	2,767	18	44	2,705	232	2,473
Israel	91%	4,315	14	105	4,196	398	3,798
Italy	92%	4,295	22	166	4,107	340	3,767
Norway (5)	88%	4,294	48	136	4,110	500	3,610
Portugal	92%	5,305	58	293	4,954	396	4,558
Singapore	95%	6,719	29	0	6,690	370	6,320
Slovenia	93%	4,676	10	35	4,631	328	4,303
Sweden	90%	4,528	34	170	4,324	445	3,879
United Arab Emirates	92%	17,208	89	232	16,887	1,321	15,566
United States	90%	4884	155	175	4554	464	4,090
Benchmarking Participants							
Abu Dhabi, UAE	92%	4,367	20	27	4,320	340	3,980
Dubai, UAE	92%	8,302	50	148	8,104	633	7,471

Students attending a sampled class at the time the sample was chosen but leaving the class before the assessment was administered were classified as "withdrawn."

Students with a disability or language barrier that prevented them from participating in the assessment were classified as "excluded."

Students not present when the assessment was administered, and not subsequently assessed in a make-up session, were classified as "absent."

In schools with 21 or fewer 4th grade students, all PIRLS students were selected to participate in ePIRLS; in larger schools, a subset of PIRLS students was randomly selected.

PIRLS 2016 Trends in Student Populations

Because a primary goal of the PIRLS 2016 assessment was to measure changes in students' reading achievement across assessment cycles, it is important to track any changes over time in population composition and coverage that might be related to student achievement. Exhibit 5.15 presents, for each country, trends across cycles (2016, 2011, 2006, and 2001) in four characteristics of the PIRLS assessment populations: number of years of formal schooling, average student age, percent of students in the national target population excluded from the assessment, and overall participation rates after using replacements. Most countries and benchmarking participants were very similar with regard to these characteristics across the four assessment cycles, although there have been changes in some countries in the age and grade structure of the assessed populations, in target population coverage, and in the exclusion rate.

The Russian Federation and Slovenia underwent structural changes in the age at which children enter schools that are reflected in their samples. In 2001, the Russian sample contained third grade students from some regions and fourth grade students from others, whereas all students were in the fourth grade by 2006. By 2011, Slovenia had completed the transition toward having all children begin school at an earlier age so that they all would have four years of primary schooling at the fourth grade instead of three years, as was the case in 2001.

National coverage of the international target population was generally comprehensive for most countries and has not changed across PIRLS assessments, with some exceptions. In 2011, Lithuania assessed only students receiving instruction in Lithuanian, and in 2016 Lithuania also assessed students receiving instruction in Russian and Polish. To ensure stable measurement of trends, the 2016 trend population for Lithuania (reported in the trend exhibits) included only students taught in Lithuanian, which represents 91 percent of the population assessed in 2016. Similarly, in 2011 Azerbaijan only tested students taught in Azerbaijani, and in 2016 Azerbaijan also tested students taught in Russian. Thus, the 2016 trend population for Azerbaijan included only students taught in Azerbaijani, representing 92 percent of the population assessed in 2016.

In general, the exclusion rates do not exceed the PIRLS 2016 guidelines of 5 percent, and have not changed very much across assessments for most countries. A few countries saw a decrease in their overall exclusion rate. From 2011 to 2016, Azerbaijan decreased its overall exclusion rate by over 5 percentage points by including students taught in Russian in the sample. Belgium (Flemish) reduced their overall exclusion rate by 5.5 percent from 2006 to 2016 by also assessing eligible students from special needs schools in 2016. Student exclusion rates were higher in 2016 than in 2011 by more than 1.5 percent in Bulgaria, Denmark, Germany, Malta, Portugal, and Singapore.

As noted by the footnotes beneath Exhibit 5.15, Austria's increased exclusions in 2016 resulted from more non-native language students within the student population, and Hong Kong SAR's increased exclusions resulted from excluding international schools and schools organized by the English Schools Foundation. Georgia excluded schools in South Ossetia and Abkhazia in both 2011 and 2016, and Singapore's increased exclusions resulted from increased enrollment in private schools. Exclusion and participation rates for South Africa in 2006 were calculated based on the entire fifth grade population in the country, whereas the exclusion rates for South Africa in 2016 were only based on students receiving instruction in English, Afrikaans, or Zulu.

Exhibit 5.15: Trends in Student Populations – PIRLS 2016

Country	Years of Formal Schooling*				Average Age at Time of Testing				Overall Exclusion Rates				Overall Participation Rates (After Replacement)			
	2016	2011	2006	2001	2016	2011	2006	2001	2016	2011	2006	2001	2016	2011	2006	2001
Australia	4	4			10.0	10.0			4.8%	4.4%			94%	93%		
Austria	4	4	4		10.3	10.3	10.3		5.6%	5.1%	5.1%		98%	98%	97%	
Azerbaijan	4	4			10.1	10.2			2.1%	7.2%			96%	100%		
Belgium (Flemish)	4		4		10.1		10.0		1.6%		7.1%		92%		91%	
Belgium (French)	4	4	4		10.0	10.1	9.9		6.0%	5.6%	3.9%		97%	82%	95%	
Bulgaria	4	4	4	4	10.8	10.7	10.9	10.9	4.3%	2.5%	6.4%	2.7%	95%	95%	94%	93%
Canada	4	4			9.9	9.9			7.5%	9.9%			86%	94%		
Chinese Taipei	4	4	4		10.1	10.2	10.1		0.9%	1.4%	2.9%		98%	99%	99%	
Czech Republic	4	4		4	10.3	10.4		10.5	3.4%	5.1%		5.0%	95%	94%		90%
Denmark	4	4	4		10.8	10.9	10.9		9.8%	7.3%	6.2%		90%	95%	96%	
England	5	5	5	5	10.3	10.3	10.3	10.2	3.7%	2.4%	2.4%	5.7%	96%	82%	92%	82%
Finland	4	4			10.8	10.8			2.4%	3.1%			96%	95%		
France	4	4	4	4	9.8	10.0	10.0	10.1	5.4%	5.2%	3.8%	5.3%	96%	97%	95%	94%
Georgia	4	4	4		9.7	10.0	10.1		3.8%	4.9%	7.3%		96%	96%	98%	
Germany	4	4	4	4	10.3	10.4	10.5	10.5	4.2%	1.9%	0.7%	1.8%	95%	95%	92%	86%
Hong Kong SAR	4	4	4	4	9.9	10.1	10.0	10.2	10.1%	11.8%	3.9%	2.8%	79%	83%	97%	97%
Hungary	4	4	4	4	10.6	10.7	10.7	10.7	4.5%	4.2%	3.7%	2.1%	97%	96%	97%	95%
Iran, Islamic Rep. of	4	4	4	4	10.2	10.2	10.2	10.4	4.1%	4.5%	3.8%	0.5%	99%	99%	99%	98%
Ireland	4	4			10.5	10.3			3.1%	2.5%			96%	95%		
Israel	4	4			10.0	10.1			24.9%	24.6%			94%	93%		
Italy	4	4	4	4	9.7	9.7	9.7	9.9	4.9%	3.7%	5.3%	2.9%	95%	95%	97%	98%
Latvia	4		4	4	10.9		11.0	11.0	7.9%		4.7%	4.6%	91%		92%	89%
Lithuania	4	4	4	4	10.8	10.7	10.7	10.9	4.2%	5.6%	5.1%	3.8%	95%	94%	92%	83%
Malta	5	5			9.7	9.8			7.9%	4.1%			96%	94%		
Morocco	4	4			10.2	10.5			1.7%	2.0%			99%	95%		
Netherlands	4	4	4	4	10.1	10.2	10.3	10.3	3.1%	3.7%	3.6%	3.7%	86%	89%	90%	87%
New Zealand	4.5 - 5.5	4.5 - 5.5	4.5 - 5.5	4.5 - 5.5	10.1	10.1	10.0	10.1	3.7%	3.3%	5.3%	3.2%	92%	93%	95%	96%
Northern Ireland	4	4			10.4	10.4			3.0%	3.5%			84%	79%		
Norway (4)	4	4	4	4	9.8	9.7	9.8	10.0	5.1%	4.2%	3.8%	2.8%	95%	71%	71%	82%
Oman	4	4			9.7	9.9			0.6%	1.5%			98%	96%		
Portugal	4	4			9.8	10.0			7.5%	2.5%			93%	93%		
Qatar	4	4			10.0	10.0			3.9%	6.2%			97%	99%		
Russian Federation	4	4	3 or 4	3 or 4	10.8	10.8	10.8	10.3	4.1%	5.3%	5.9%	6.6%	98%	98%	97%	97%
Saudi Arabia	4	4			9.9	10.0			2.3%	1.6%			96%	98%		

Exhibit 5.15: Trends in Student Populations – PIRLS 2016 (Continued)

Country	Years of Formal Schooling*				Average Age at Time of Testing				Overall Exclusion Rates				Overall Participation Rates (After Replacement)			
	2016	2011	2006	2001	2016	2011	2006	2001	2016	2011	2006	2001	2016	2011	2006	2001
Singapore	4	4	4	4	10.4	10.4	10.4	10.1	11.1%	6.3%	0.9%	0.1%	97%	96%	95%	98%
Slovak Republic	4	4	4	4	10.4	10.4	10.4	10.3	4.8%	4.6%	3.6%	2.0%	97%	96%	94%	96%
Slovenia	4	4	3 or 4	3	9.9	9.9	9.9	9.8	2.4%	2.6%	0.8%	0.3%	90%	94%	93%	94%
South Africa	4	4			10.6	10.5			2.5%	3.0%			94%	95%		
Spain	4	4	4		9.9	9.9	9.9		4.8%	5.4%	5.3%		97%	96%	97%	
Sweden	4	4	4	4	10.7	10.7	10.9	10.8	5.2%	4.1%	3.9%	5.0%	95%	91%	96%	92%
Trinidad and Tobago	5	5	5		10.2	10.3	10.1		1.3%	0.9%	0.7%		96%	95%	94%	
United Arab Emirates	4	4			9.8	9.8			3.3%	3.3%			95%	97%		
United States	4	4	4	4	10.1	10.2	10.1	10.2	4.8%	7.2%	5.9%	5.3%	86%	81%	82%	83%
Benchmarking Participants																
Ontario, Canada	4	4	4	4	9.8	9.9	9.8	9.9	4.1%	7.9%	8.3%	6.6%	93%	95%	87%	92%
Quebec, Canada	4	4	4	4	10.1	10.1	10.1	10.2	5.1%	3.7%	3.6%	3.3%	64%	92%	81%	89%
Eng/Afr/Zulu - RSA (5)	5		5		11.6		11.7		1.1%		4.3%		86%		88%	
Andalusia, Spain	4	4			9.8	9.9			4.2%	5.1%			96%	96%		
Abu Dhabi, UAE	4	4			9.7	9.7			3.9%	2.7%			96%	96%		
Dubai, UAE	4	4			9.9	9.9			3.2%	5.1%			95%	94%		

* Represents years of schooling counting from the first year of ISCED Level 1.

An empty cell indicates a country did not participate in that year's assessment or did not have comparable data.

Trend results for Azerbaijan do not include students taught in Russian. Trend results for Lithuania do not include students taught in Polish or Russian.

Austria's increased exclusions in 2016 resulted from more non-native language speakers, probably due to the refugee crisis in Europe.

Canada's decreased exclusions in 2016 resulted from provinces formerly reported as exclusions to be considered not covered by the target population.

Georgian schools in South Ossetia and Abkhazia were excluded in 2011 and 2016 due to lack of access and absence of official statistics. Abkhazia refugee schools in other territories of Georgia were included in the sample frame.

Hong Kong SAR's increased exclusions in 2011 and 2016 resulted from excluding international schools and schools organized by the English Schools Foundation. These schools do not follow Hong Kong's central curriculum and medium of instruction.

Singapore's increased exclusions in 2016 resulted from increased enrollment in private schools, which predominantly serve international students and are different from public schools in many respects (e.g., different language of instruction and calendar year).

Republic of South Africa (RSA) tested 5th grade students receiving instruction in English (Eng), Afrikaans (Afr) and Zulu. Exclusion and participation rates from 2006 are for the entire country of South Africa.

Appendix 5A: Characteristics of National Samples

Australia

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and very remote schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by state or territory (8)
- Implicit stratification by geographic location (metropolitan, provincial, remote), school type (Catholic, government, independent), and socioeconomic index (low socioeconomic status, high socioeconomic status)
- Prior to class sampling within schools, all indigenous students were grouped into a single classroom and were selected with certainty. The other classroom in the school was sampled using the standard procedure.
- Schools were oversampled at the state/territory level

Allocation of School Sample in Australia

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Australian Capital Territory	30	0	29	1	0	0	0
New South Wales	45	0	42	2	1	0	0
Northern Territory	15	0	15	0	0	0	0
Queensland	45	0	45	0	0	0	0
South Australia	41	0	41	0	0	0	0
Tasmania	27	0	27	0	0	0	0
Victoria	44	0	43	1	0	0	0
Western Australia	39	0	39	0	0	0	0
Total	**286**	**0**	**281**	**4**	**1**	**0**	**0**

Austria

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3) and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers
- Exclusion rates are higher than usual because of more non-native language speakers in classes. This higher proportion of non-native language speakers is probably due to the refugee crisis in Europe.

Sample Design

- Explicit stratification by region (9)
- No implicit stratification
- Sampled two classrooms per school whenever possible

Allocation of School Sample in Austria

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Burgenland	8	1	7	0	0	0	0
Kärnten	10	0	10	0	0	0	0
Niederösterreich	28	0	28	0	0	0	0
Oberösterreich	26	0	26	0	0	0	0
Salzburg	10	0	10	0	0	0	0
Steiermark	20	0	20	0	0	0	0
Tirol	12	0	12	0	0	0	0
Vorarlberg	8	0	8	0	0	0	0
Wien	30	1	29	0	0	0	0
Total	**152**	**2**	**150**	**0**	**0**	**0**	**0**

Azerbaijan

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and schools with English and Georgian instructional language
- No within-school exclusions
- Exclusion rates are biased downward due to exclusion of Armenian community schools in the Nagorno-Karabakh conflict zone and international schools for which no statistics were available

Sample Design

- Explicit stratification by language of instruction (Azerbaijani only, Russian or Russian/ Azerbaijani), urbanization (urban, rural) within Azerbaijani only strata, and city (Baku, other) within urban stratum
- No implicit stratification
- Sampled two classrooms in schools with four or more classrooms

Allocation of School Sample in Azerbaijan

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Azerbaijani - Urban - Baku	24	0	24	0	0	0	0
Azerbaijani - Urban - Other cities	38	0	38	0	0	0	0
Azerbaijani - Rural	68	0	68	0	0	0	0
Russian or Russian/ Azerbaijani	40	0	40	0	0	0	0
Total	**170**	**0**	**170**	**0**	**0**	**0**	**0**

Bahrain

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), special needs schools, students taught in French, and students taught in Japanese
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by governorate (5) and gender (girls, boys) within public schools
- No implicit stratification
- Sampled one classroom per school
- All schools were selected
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates, when all classes within school were sampled

Allocation of School Sample in Bahrain

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public Muharraq - Girls	10	0	10	0	0	0	0
Public Muharraq - Boys	11	0	11	0	0	0	0
Public Capital - Girls	19	0	19	0	0	0	0
Public Capital - Boys	21	0	21	0	0	0	0
Public Northern - Girls	21	0	21	0	0	0	0
Public Northern - Boys	17	0	17	0	0	0	0
Public Southern - Girls	11	0	11	0	0	0	0
Public Southern - Boys	11	0	11	0	0	0	0
Private	63	1	61	0	0	1	0
Total	**184**	**1**	**182**	**0**	**0**	**1**	**0**

Belgium (Flemish)

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5) and French schools
- Within-school exclusions consisted of students with intellectual disabilities and non-native language speakers

Sample Design

- Explicit stratification by region (6), socioeconomic status (4), school type (official, private), and a stratum of eligible special education schools
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 37)
- Field Test and Main Data Collection samples were selected separately. PIRLS Field Test sample was selected simultaneously with the TIMSS 2015 Main Data Collection sample to avoid overlap. PIRLS Main Data Collection sample was selected using the Chowdhury method to minimize overlap with both PIRLS Field Test sample and TIMSS 2015 Main Data Collection sample.

Allocation of School Sample in Belgium (Flemish)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Antwerpen - Official - Low SES	9	0	6	2	0	1	0
Antwerpen - Private - Low SES	8	0	4	3	1	0	0
Antwerpen - High SES	8	0	7	1	0	0	0
Antwerpen - Med-High SES	8	1	7	0	0	0	0
Antwerpen - Med-Low SES	8	0	7	0	0	1	0
Brussels Hoofdstedelijk Gewest - Low SES	8	0	6	1	0	1	0
Limburg - Higher SES	10	0	6	3	1	0	0
Limburg - Lower SES	10	0	5	3	0	2	0
Oost-Vlaanderen - High SES	8	0	7	0	0	1	0
Oost-Vlaanderen - Med-High SES	7	0	6	0	0	1	0
Oost-Vlaanderen - Med-Low SES	10	0	10	0	0	0	0
Oost-Vlaanderen - Low SES	8	0	5	3	0	0	0
Vlaams-Brabant - Higher SES	12	0	8	1	2	1	0
Vlaams-Brabant - Lower SES	12	0	12	0	0	0	0
West-Vlaanderen - High SES	8	0	8	0	0	0	0
West-Vlaanderen - Med-High SES	7	0	7	0	0	0	0
West-Vlaanderen - Lower SES	9	0	8	1	0	0	0
Special Education schools	10	2	5	1	1	1	0
Total	**160**	**3**	**124**	**19**	**5**	**9**	**0**

Belgium (French)

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5) and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school network (public at state level, public at local level, private) and socioeconomic status (4)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 47)

Allocation of School Sample in Belgium (French)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public at state level - 1st and 2nd SES quartiles	8	0	8	0	0	0	0
Public at state level - 3rd and 4th SES quartiles	8	0	8	0	0	0	0
Public at local level - 1st SES quartile	18	0	18	0	0	0	0
Public at local level - 2nd SES quartile	16	0	15	1	0	0	0
Public at local level - 3rd SES quartile	20	0	20	0	0	0	0
Public at local level - 4th SES quartile	22	0	22	0	0	0	0
Private sectarian - 1st SES quartile	14	0	13	1	0	0	0
Private sectarian - 2nd SES quartile	14	0	13	1	0	0	0
Private sectarian - 3rd SES quartile	20	0	18	2	0	0	0
Private sectarian - 4th SES quartile	18	0	17	1	0	0	0
Total	**158**	**0**	**152**	**6**	**0**	**0**	**0**

Bulgaria

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5) and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (elementary, basic, general) and urbanization (capital, large cities, other)
- Implicit stratification by urbanization (city, village) within the basic schools found outside the larger cities
- Sampled two classrooms in large schools (measure of size > 69)
- The school sample was selected by controlling for the overlap with the TIMSS 2015 sample using the Chowdhury approach

Allocation of School Sample in Bulgaria

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Elementary - Capital and Large Cities	8	0	8	0	0	0	0
Elementary - Others	8	0	8	0	0	0	0
Basic - Capital	10	1	9	0	0	0	0
Basic - Large Cities	29	0	29	0	0	0	0
Basic - Others	44	0	44	0	0	0	0
General - Capital	14	0	14	0	0	0	0
General - Large Cities	17	0	17	0	0	0	0
General - Others	24	0	24	0	0	0	0
Total	**154**	**1**	**153**	**0**	**0**	**0**	**0**

Canada

Coverage and Exclusions

- Coverage is 96.9 percent. Coverage in Canada is restricted to students from the provinces of Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland, Ontario, Quebec, and Saskatchewan.
- School-level exclusions consisted of very small schools (measure of size < 4 in Manitoba and Saskatchewan; measure of size < 6 in Alberta, Newfoundland, and Ontario; measure of size < 9 in British Columbia; and measure of size < 10 in Quebec); special needs schools, First Nations, French first language (in Newfoundland); home schooled, institutional, and private schools as well as public special schools (in Manitoba); international schools, non-ministry, and special status schools (in Quebec); and distance learning and not funded schools (in British Colombia)
- For ePIRLS, coverage is 74 percent. Coverage in Canada is restricted to students from the provinces of British Columbia, Newfoundland, Ontario, and Quebec.
- For ePIRLS, school-level exclusions consisted of very small schools (measure of size < 4 in Manitoba and Saskatchewan, measure of size < 6 in Alberta, Newfoundland and Ontario, measure of size < 9 in British Columbia, and measure of size < 10 in Quebec); special needs schools, First Nations, French first language (in Newfoundland); international schools, non-ministry, and special status schools (in Quebec); and distance learning and not funded schools (in British Colombia)
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by province (8). Within the province of British Columbia, explicit stratification was done by school language (English, French) and school type within English schools (English only, immersion, dual track). Within the province of Alberta, explicit stratification was done by school system (French, English) and school type (immersion, regular). Within the province of Ontario, explicit stratification was done by school type (private, Catholic, public) and language (English, French) within Catholic and public schools. Within Quebec, explicit stratification was done by school type (public, private) and language (French, English). Within the province of New Brunswick, explicit stratification was done by school language (English, French)
- Implicit stratification by region (4) in public and Catholic explicit strata within Ontario

- Sampled two classrooms in large schools for Quebec and Ontario (measure of size > 80), as well as in Alberta French schools. All classrooms selected in British Columbia French schools.
- The PIRLS school sample was selected by controlling for the overlap with the TIMSS 2015 Grade 4 sample using the Chowdhury approach
- All French schools in British Columbia were selected
- For ePIRLS, only a subsample of PIRLS schools was randomly selected in Quebec. School weights were adjusted accordingly.
- In British Columbia French schools stratum, schools or classes were used as variance estimation strata and half classes were used as jackknife replicates

Allocation of School Sample in Canada - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Newfoundland	130	0	128	0	0	2	0
New Brunswick - English	136	6	130	0	0	0	0
New Brunswick - French	66	0	66	0	0	0	0
Quebec - English - Private	8	1	7	0	0	0	0
Quebec - English - Public	42	0	39	0	0	3	0
Quebec - French - Private	8	0	8	0	0	0	0
Quebec - French - Public	118	1	35	25	13	44	0
Ontario - Private	8	0	0	1	0	7	0
Ontario - English - Catholic	30	0	30	0	0	0	0
Ontario - English - Public	80	2	77	1	0	0	0
Ontario - French - Catholic & Public	80	0	79	0	0	1	0
Manitoba	8	1	7	0	0	0	0
Saskatchewan	8	0	6	2	0	0	0
British Columbia - English System - English	106	1	104	0	0	1	0
British Columbia - English System - Immersion	8	0	8	0	0	0	0
British Columbia - English System - Dual Track	18	1	17	0	0	0	0
British Columbia - French System	17	0	17	0	0	0	0
Alberta - English System - Private	6	1	4	1	0	0	2
Alberta - English System - Public	17	1	10	1	1	4	1
Alberta - English System - French Immersion - Private	6	1	3	1	0	1	1
Alberta - English System - French Immersion - Public	90	0	75	7	1	7	2
Alberta - French System - Public	24	0	22	0	0	2	0
Total	**1014**	**16**	**872**	**39**	**15**	**72**	**6**

Allocation of School Sample in Canada - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Newfoundland	130	0	127	0	0	3	0
Quebec - English - Private	2	0	2	0	0	0	0
Quebec - English - Public	2	0	2	0	0	0	0
Quebec - French - Private	2	0	2	0	0	0	0
Quebec - French - Public	24	0	8	4	1	11	0
Ontario - Private	8	0	0	1	0	7	0
Ontario - English - Catholic	30	0	30	0	0	0	0
Ontario - English - Public	80	2	75	1	0	2	0
Ontario - French - Catholic & Public	80	0	77	0	0	3	0
British Columbia - English System - English	106	1	102	0	0	3	0
British Columbia - English System - Immersion	8	0	8	0	0	0	0
British Columbia - English System - Dual Track	18	1	17	0	0	0	0
British Columbia - French System	17	0	17	0	0	0	0
Total	**507**	**4**	**467**	**6**	**1**	**29**	**0**

Chile

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and geographically inaccessible schools
- Within-school exclusions consisted of students with intellectual disabilities

Sample Design

- Explicit stratification by school type (public, private subsidized, private paid), urbanization (rural, urban) within public schools and school size (up to 40 students, 41-80 students, more than 80 students) within public and private subsidized schools
- No implicit stratification
- Sampled one classroom
- The school sample for PIRLS was selected by controlling for the overlap with the ICCS sample using the Chowdhury approach

Allocation of School Sample in Chile

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public - Urban - Up to 40 students	14	0	14	0	0	0	0
Public - Urban - 41 to 80 students	16	0	16	0	0	0	0
Public - Urban - 80 or more students	8	0	8	0	0	0	0
Public - Rural	9	0	9	0	0	0	0
Private subsidized - Up to 40 students	20	0	17	3	0	0	0
Private subsidized - 41 to 80 students	24	0	22	1	1	0	0
Private subsidized - 80 or more students	24	0	21	3	0	0	0
Private	39	0	32	7	0	0	0
Total	**154**	**0**	**139**	**14**	**1**	**0**	**0**

Chinese Taipei

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (north, middle, south, east and isolated islands). East and isolated islands were grouped together.
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 289)

Allocation of School Sample in Chinese Taipei - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
North	66	0	66	0	0	0	0
Middle	38	0	38	0	0	0	0
South	38	0	38	0	0	0	0
East & Isolated Islands	8	0	8	0	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**	**0**

Allocation of School Sample in Chinese Taipei - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
North	66	0	66	0	0	0	0
Middle	38	0	38	0	0	0	0
South	38	0	38	0	0	0	0
East & Isolated Islands	8	0	8	0	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**	**0**

Czech Republic

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), special needs schools, and Polish instructional language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (14)
- No implicit stratification
- Sampled two classrooms whenever possible

Allocation of School Sample in Czech Republic

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Praha	17	0	17	0	0	0	0
Středočeský	20	0	20	0	0	0	0
Jihočeský	9	0	9	0	0	0	0
Plzeňský	8	0	8	0	0	0	0
Karlovarský	8	0	8	0	0	0	0
Ústecký	12	0	12	0	0	0	0
Liberecký	8	0	8	0	0	0	0
Královéhradecký	8	0	8	0	0	0	0
Pardubický	8	0	8	0	0	0	0
Vysočina	8	0	8	0	0	0	0
Jihomoravský	17	0	17	0	0	0	0
Olomoucký	9	0	9	0	0	0	0
Zlínský	8	0	8	0	0	0	0
Moravskoslezský	17	0	17	0	0	0	0
Total	**157**	**0**	**157**	**0**	**0**	**0**	**0**

Denmark

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, daycare and rehabilitation home schools as well as German, English, and Rudolf Steiner schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private)
- No implicit stratification
- Sampled one classroom per school

Allocation of School Sample in Denmark - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public	171	7	154	8	0	2	0
Private	27	0	16	6	1	4	0
Total	**198**	**7**	**170**	**14**	**1**	**6**	**0**

Allocation of School Sample in Denmark - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public	171	7	124	5	0	35	0
Private	27	0	8	4	1	14	0
Total	**198**	**7**	**132**	**9**	**1**	**49**	**0**

Egypt

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), schools in Matrouh, and schools in North Sinai
- No within-school exclusions

Sample design

- Explicit stratification by region (Capital, North, South) and school type (government, private)
- Implicit stratification by urbanization (urban, rural) within government schools strata
- Sampled one classroom per school

Allocation of School Sample in Egypt

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Capital - Government	30	0	30	0	0	0	0
Capital - Private	10	0	10	0	0	0	0
North - Government	60	0	60	0	0	0	0
North - Private	8	0	8	0	0	0	0
South - Government	44	0	44	0	0	0	0
South - Private	8	0	8	0	0	0	0
Total	**160**	**0**	**160**	**0**	**0**	**0**	**0**

England

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 8), special needs schools, and pupil referral units
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (state-funded, private) and attainment level (5)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 99)
- The Field Test and Main Data Collection PIRLS samples were selected separately. The PIRLS Main Data Collection sample was selected by controlling for the overlap with the TIMSS 2015 samples and with the PIRLS Field Test sample using the Chowdhury approach.

Allocation of School Sample in England

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
State-funded - Low attainment level	26	0	25	1	0	0	0
State-funded - Low to Mid attainment level	34	0	34	0	0	0	0
State-funded - Mid and missing attainment level	34	0	33	1	0	0	0
State-funded - Mid to High attainment level	35	0	35	0	0	0	0
State-funded - High attainment level	30	0	30	0	0	0	0
Private	12	1	11	0	0	0	0
Total	**171**	**1**	**168**	**2**	**0**	**0**	**0**

Finland

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools and schools with instructional languages other than Finnish or Swedish
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by major region (Helsinki and Uusimaa, southern, western, northern) and urbanization (urban and semi-urban, rural) within Finnish schools. Swedish speaking schools are in a separate explicit stratum.
- No implicit stratification
- Sampled two classrooms per school
- The PIRLS samples were selected by controlling for the overlap with the TIMSS 2015 Main Data Collection sample using the Chowdhury approach

Allocation of School Sample in Finland

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Helsinki & Uusimaa	40	0	39	0	0	1	0
Southern - Urban & Semi-Urban	26	3	22	1	0	0	0
Southern - Rural	8	2	6	0	0	0	0
Western - Urban & Semi-Urban	32	1	31	0	0	0	0
Western - Rural	8	0	8	0	0	0	0
Northern & Eastern - Urban & Semi-Urban	26	0	25	1	0	0	0
Northern & Eastern - Rural	10	1	9	0	0	0	0
Swedish speaking	9	0	9	0	0	0	0
Total	**159**	**7**	**149**	**2**	**0**	**1**	**0**

France

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 3), overseas territories, Reunion and Mayotte Islands, Guyana (Southern Hemisphere), private schools without contract, specialized schools, and French schools in foreign countries
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public-other, public-priority education zone, private)
- No implicit stratification
- Sampled two classrooms per school
- PIRLS 2016 samples and TIMSS 2015 samples were selected simultaneously to avoid overlap between the two studies

Allocation of School Sample in France

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public-other	100	2	98	0	0	0	0
Public-priority education zone	44	1	42	1	0	0	0
Private	22	0	21	1	0	0	0
Total	**166**	**3**	**161**	**2**	**0**	**0**	**0**

Georgia

Coverage and Exclusions

- Coverage is 95.9 percent. Coverage in Georgia is restricted to students taught in Georgian and Azerbaijani.
- School-level exclusions consisted of very small schools (measure of size < 3) and foreign instructional language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by language taught in school (Georgian, Azerbaijani), teacher certification (certified, non-certified), urbanization (urban, rural), and school type (public, private)
- No implicit stratification
- Sampled two classrooms in Georgian schools with certified teachers
- The Field Test and Main Data Collection PIRLS samples were selected sequentially. The PIRLS Main Data Collection sample was selected by controlling for the overlap with the PIRLS Field Test sample using the Chowdhury approach.
- Oversampled Azerbaijani schools as well as public schools with certified teachers in order to get better estimates
- Class group option was used in bilingual schools as well as in schools with certified teachers

Allocation of School Sample in Georgia - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Georgian - Certified - Urban - Public	71	0	71	0	0	0	0
Georgian - Certified - Rural - Public	16	0	16	0	0	0	0
Georgian - Certified - Private	8	0	8	0	0	0	0
Georgian - Non-certified - Urban - Public	33	0	33	0	0	0	0
Georgian - Non-certified - Rural - Public	35	0	34	1	0	0	0
Georgian - Non-certified - Private	8	0	6	0	1	1	0
Azeri	30	0	30	0	0	0	0
Total	**201**	**0**	**198**	**1**	**1**	**1**	**0**

Allocation of School Sample in Georgia - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Georgian - Certified - Urban - Public	71	0	70	0	0	1	0
Georgian - Certified - Rural - Public	16	0	16	0	0	0	0
Georgian - Certified - Private	8	0	8	0	0	0	0
Georgian - Non-certified - Urban - Public	33	0	33	0	0	0	0
Georgian - Non-certified - Rural - Public	35	0	34	1	0	0	0
Georgian - Non-certified - Private	8	0	6	0	1	1	0
Azeri	30	0	30	0	0	0	0
Total	**201**	**0**	**197**	**1**	**1**	**2**	**0**

Germany

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by percentage of immigrants in school (very low, low, medium, high). A separate stratum was created for the special needs schools (SEN).
- No implicit stratification
- Sampled one classroom per school

Allocation of School Sample in Germany

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Regular - Very low	62	1	57	2	1	1	0
Regular - Low	94	0	94	0	0	0	0
Regular - Medium	28	0	27	1	0	0	0
Regular - High	16	0	16	0	0	0	0
Special needs schools	10	0	10	0	0	0	0
Total	**210**	**1**	**204**	**3**	**1**	**1**	**0**

Hong Kong SAR

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools and international schools
- Within-school exclusions consisted of students with intellectual disabilities and non-native language speakers

Sample Design

- Explicit stratification by school gender (single gender, co-educational) and school type (4) within co-educational strata
- No implicit stratification
- Sampled two classrooms in large co-educational aided schools with six or more classrooms

Allocation of School Sample in Hong Kong SAR

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Single gender	8	0	6	1	1	0	0
Co-educational - Aided	120	1	89	15	5	10	0
Co-educational - Direct subsidy	8	0	7	0	0	1	0
Co-educational - Government	8	0	8	0	0	0	0
Co-educational - Private	8	0	4	2	0	2	0
Total	**152**	**1**	**114**	**18**	**6**	**13**	**0**

Hungary

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and students taught in foreign language
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by community type (capital and county town, town, rural area) and national assessment reading score (low, medium, high, missing)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 55)

Allocation of School Sample in Hungary

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Capital and County Town - Low or Medium score	16	0	16	0	0	0	0
Capital and County Town - High score	30	0	29	1	0	0	0
Capital and County Town - Missing score	8	1	6	1	0	0	0
Town - Low score	14	0	14	0	0	0	0
Town - Medium score	20	1	19	0	0	0	0
Town - High score	14	1	12	1	0	0	0
Town - Missing score	12	0	12	0	0	0	0
Rural Area - Low score	14	0	14	0	0	0	0
Rural Area - Medium score	10	0	10	0	0	0	0
Rural Area - High score	8	0	8	0	0	0	0
Rural Area - Missing score	8	2	6	0	0	0	0
Total	**154**	**5**	**146**	**3**	**0**	**0**	**0**

Iran, Islamic Rep. of

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and geographically inaccessible schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), gender (mixed, other), region group (1, 2, 3), province or grouped provinces (6), and gender (boys, girls) within "other" gender public schools
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 119)
- The Field Test and Main Data Collection PIRLS samples were selected separately
- PIRLS and PIRLS Literacy booklets were rotated within classes

Allocation of School Sample in Iran, Islamic Rep. of

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Private	16	0	16	0	0	0	0
Public - Mixed - Region group 1	10	0	10	0	0	0	0
Public - Mixed - Region group 2	8	0	8	0	0	0	0
Public - Mixed - Region group 3	8	1	7	0	0	0	0
Public - Other - Region group 1 - All others provinces - Boys	12	0	12	0	0	0	0
Public - Other - Region group 1 - All others provinces - Girls	12	0	12	0	0	0	0
Public - Other - Region group 1 - Khozestan - Boys	14	0	14	0	0	0	0
Public - Other - Region group 1 - Khozestan - Girls	14	0	14	0	0	0	0
Public - Other - Region group 2 - All others provinces - Boys	10	0	10	0	0	0	0
Public - Other - Region group 2 - All others provinces - Girls	10	0	10	0	0	0	0
Public - Other - Region group 2 - Razavi Khorasan - Boys	14	0	14	0	0	0	0
Public - Other - Region group 2 - Razavi Khorasan - Girls	14	0	14	0	0	0	0
Public - Other - Region group 2 - Tehran Province - Boys	14	0	14	0	0	0	0
Public - Other - Region group 2 - Tehran Province - Girls	14	0	14	0	0	0	0

Allocation of School Sample in Iran, Islamic Rep. of (Continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public - Other - Region group 3 - All others provinces - Boys	10	0	10	0	0	0	0
Public - Other - Region group 3 - All others provinces - Girls	10	0	10	0	0	0	0
Public - Other - Region group 3 - Esfahan - Boys	14	1	13	0	0	0	0
Public - Other - Region group 3 - Esfahan - Girls	14	0	14	0	0	0	0
Public - Other - Region group 3 - Fars - Boys	14	1	13	0	0	0	0
Public - Other - Region group 3 - Fars - Girls	14	0	14	0	0	0	0
Public - Other - Region group 3 - Tehran City - Boys	14	0	14	0	0	0	0
Public - Other - Region group 3 - Tehran City - Girls	14	0	14	0	0	0	0
Total	**274**	**3**	**271**	**0**	**0**	**0**	**0**

Ireland

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, and non-aided private schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school level socioeconomic status DEIS (non-DEIS, rural, urban band 1, urban band 2), school type (ordinary, Gaeltacht, Gaelscoil), and gender (boys, girls, mixed)
- No implicit stratification
- Sampled two classrooms per school
- For ePIRLS, students were subsampled within classes and students weights were adjusted accordingly

Allocation of School Sample in Ireland - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Gaelscoil	10	0	10	0	0	0	0
Gaeltacht	8	0	8	0	0	0	0
Non-DEIS - Ordinary - Boys	12	1	11	0	0	0	0
Non-DEIS - Ordinary - Girls	12	0	12	0	0	0	0
Non-DEIS - Ordinary - Mixed	80	0	80	0	0	0	0
Rural - Ordinary	8	1	7	0	0	0	0
Urban Band 1 - Ordinary	10	0	10	0	0	0	0
Urban Band 2 - Ordinary	10	0	10	0	0	0	0
Total	**150**	**2**	**148**	**0**	**0**	**0**	**0**

Allocation of School Sample in Ireland - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Gaelscoil	10	0	10	0	0	0	0
Gaeltacht	8	0	8	0	0	0	0
Non-DEIS - Ordinary - Boys	12	1	11	0	0	0	0
Non-DEIS - Ordinary - Girls	12	0	12	0	0	0	0
Non-DEIS - Ordinary - Mixed	80	0	80	0	0	0	0
Rural - Ordinary	8	1	6	0	0	1	0
Urban Band 1 - Ordinary	10	0	10	0	0	0	0
Urban Band 2 - Ordinary	10	0	10	0	0	0	0
Total	**150**	**2**	**147**	**0**	**0**	**1**	**0**

Israel

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, schools teaching in English or French, and Ultra-Orthodox schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school sector (Hebrew-Secular, Hebrew-Religious, Arabic), socioeconomic status (high, medium, low) and subgroups within Arab sector (Arab, Bedouin, Druze)
- No implicit stratification
- Sampled one classroom per school

Allocation of School Sample in Israel - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Hebrew-Secular - High SES	42	0	41	0	0	1	0
Hebrew-Secular - Medium SES	26	0	25	1	0	0	0
Hebrew-Secular - Low SES	10	0	10	0	0	0	0
Hebrew-Religious - High SES	10	0	9	1	0	0	0
Hebrew-Religious - Medium SES	16	0	16	0	0	0	0
Hebrew-Religious - Low SES	8	0	8	0	0	0	0
Arabic-Arab - Medium SES	10	0	10	0	0	0	0
Arabic-Arab - Low SES	18	0	18	0	0	0	0
Arabic-Bedouin	12	0	12	0	0	0	0
Arabic-Druze	8	0	8	0	0	0	0
Total	**160**	**0**	**157**	**2**	**0**	**1**	**0**

Allocation of School Sample in Israel - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Hebrew-Secular - High SES	42	0	41	0	0	1	0
Hebrew-Secular - Medium SES	26	0	25	1	0	0	0
Hebrew-Secular - Low SES	10	0	10	0	0	0	0
Hebrew-Religious - High SES	10	0	9	1	0	0	0
Hebrew-Religious - Medium SES	16	0	15	0	0	1	0
Hebrew-Religious - Low SES	8	0	7	0	0	1	0
Arabic-Arab - Medium SES	10	0	10	0	0	0	0
Arabic-Arab - Low SES	18	0	18	0	0	0	0
Arabic-Bedouin	12	0	12	0	0	0	0
Arabic-Druze	8	0	8	0	0	0	0
Total	**160**	**0**	**155**	**2**	**0**	**3**	**0**

Italy

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of Slovenian, Ladin, and German instructional language schools
- Within-school exclusions consisted of students with functional disabilities and non-native language speakers

Sample Design

- Explicit stratification by school type (private, public) and region (center, south and islands, north east, north west, south) within public schools
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 109)
- The Field Test and Main Data Collection PIRLS samples were selected separately. The PIRLS Main Data Collection sample was selected by controlling for the overlap with the TIMSS 2015 and PIRLS Field Test samples using the Chowdhury approach.

Allocation of School Sample in Italy - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Private	10	0	8	2	0	0	0
Public - Center	28	0	24	4	0	0	0
Public - South and Islands	22	0	20	2	0	0	0
Public - North East	26	0	21	4	0	1	0
Public - North West	36	0	34	2	0	0	0
Public - South	28	0	27	1	0	0	0
Total	**150**	**0**	**134**	**15**	**0**	**1**	**0**

Allocation of School Sample in Italy - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Private	10	0	8	2	0	0	0
Public - Center	28	0	24	4	0	0	0
Public - South and Islands	22	0	19	2	0	1	0
Public - North East	26	0	21	4	0	1	0
Public - North West	36	0	34	2	0	0	0
Public - South	28	0	27	1	0	0	0
Total	**150**	**0**	**133**	**15**	**0**	**2**	**0**

Kazakhstan

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, and languages other than Kazakh and Russian
- No within-school exclusions

Sample Design

- Explicit stratification by region (4), language (Kazakh, Russian, both languages) and urbanization (urban, rural)
- No implicit stratification
- Sampled two classrooms in schools with both Kazakh and Russian languages of instruction
- Class group option was used in bilingual schools

Allocation of School Sample in Kazakhstan

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Region A - Kazakh - Urban	10	0	10	0	0	0	0
Region A - Kazakh - Rural	18	1	17	0	0	0	0
Region A - Russian	8	0	7	1	0	0	0
Region A - Both Kazakh and Russian	12	0	12	0	0	0	0
Region B - Kazakh - Urban	8	0	8	0	0	0	0
Region B - Kazakh - Rural	10	0	10	0	0	0	0
Region B - Both Kazakh and Russian - Urban	8	0	8	0	0	0	0
Region B - Both Kazakh and Russian - Rural	10	0	10	0	0	0	0
Region B and C - Russian/Other	14	0	14	0	0	0	0
Region C - Kazakh - Urban	12	0	12	0	0	0	0
Region C - Kazakh - Rural	8	1	7	0	0	0	0
Region C - Both Kazakh and Russian - Urban	16	0	16	0	0	0	0
Region C - Both Kazakh and Russian - Rural	8	0	8	0	0	0	0
Region D - Kazakh - Urban	8	0	8	0	0	0	0
Region D - Kazakh - Rural	8	0	8	0	0	0	0
Region D - Russian	8	0	8	0	0	0	0
Region D - Both Kazakh and Russian	8	0	8	0	0	0	0
Total	**174**	**2**	**171**	**1**	**0**	**0**	**0**

Kuwait

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools and minority language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private), region (6), and gender (male, female) within public schools, and language (Arabic, foreign, bilingual) within private schools
- No implicit stratification
- Sampled two classrooms in private bilingual schools
- The PIRLS samples were selected simultaneously with the TIMSS Main Data Collection to avoid overlap
- All private bilingual were sampled for PIRLS

Allocation of School Sample in Kuwait

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public - Asema – Female	10	1	9	0	0	0	0
Public - Asema – Male	11	0	11	0	0	0	0
Public - Hawally – Female	8	0	8	0	0	0	0
Public - Hawally - Male	8	0	8	0	0	0	0
Public - Farwaniya - Female	11	0	11	0	0	0	0
Public - Farwaniya - Male	12	0	12	0	0	0	0
Public - Ahmadi - Female	12	0	12	0	0	0	0
Public - Ahmadi - Male	13	1	12	0	0	0	0
Public - Jahra - Female	10	0	10	0	0	0	0
Public - Jahra - Male	10	0	10	0	0	0	0
Public - Mubarak Alkabeer - Female	7	0	7	0	0	0	1
Public - Mubarak Alkabeer - Male	6	0	6	0	0	0	1
Private - Arabic	18	2	16	0	0	0	0
Private - Foreign	29	0	27	0	0	2	0
Private - Bilingual	20	0	18	0	0	2	0
Total	**185**	**4**	**177**	**0**	**0**	**4**	**2**

Latvia

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools and schools with instructional language other than Latvian or Russian
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school level (Grade 4 only, Grade 4 and 8), urbanization (Riga, city, town and rural area), language (Latvian, Russian), and school type (gymnasium-secondary, basic-beginners) within town and rural area Latvian schools
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 59)
- Did not participate in the Field Test. The PIRLS Main data Collection sample was selected simultaneously with the 2016 ICCS Main Data Collection sample to avoid overlap.
- Class group option was used in bilingual schools

Allocation of School Sample in Latvia

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Grade 4 only - Riga	6	1	5	0	0	0	0
Grade 4 only - City	8	0	8	0	0	0	0
Grade 4 only - Town-Rural	8	0	8	0	0	0	0
Grade 4 & Grade 8 - Riga - Latvian	22	0	22	0	0	0	0
Grade 4 & Grade 8 - Riga - Russian	24	0	21	1	0	2	0
Grade 4 & Grade 8 - City - Latvian	10	0	10	0	0	0	0
Grade 4 & Grade 8 - City - Russian	12	0	11	1	0	0	0
Grade 4 & Grade 8 - Town-Rural - Latvian - Gymnasium-Secondary	34	0	31	1	0	2	0
Grade 4 & Grade 8 - Town-Rural - Latvian - Basic-Beginners	24	0	23	0	1	0	0
Grade 4 & Grade 8 - Town-Rural - Russian	8	0	7	1	0	0	0
Total	**156**	**1**	**146**	**4**	**1**	**4**	**0**

Lithuania

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and other language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by language (Lithuanian, Russian, Polish, mixed) and urbanization within Lithuanian schools (4)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 99) and in bilingual schools
- The Field Test and Main data Collection PIRLS samples were selected sequentially
- Class group option was used in bilingual schools

Allocation of School Sample in Lithuania

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Lithuanian - Capital	22	0	22	0	0	0	0
Lithuanian - Other Major City	33	0	33	0	0	0	0
Lithuanian - City	52	0	52	0	0	0	0
Lithuanian - Small City or Village	29	0	29	0	0	0	0
Russian	20	0	20	0	0	0	0
Polish	19	0	19	0	0	0	1
Mixed	20	0	20	0	0	0	0
Total	**195**	**0**	**195**	**0**	**0**	**0**	**1**

Macao SAR

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of international schools
- Within-school exclusions consisted of students with functional disabilities and non-native language speakers

Sample Design

- All schools were sampled and therefore no explicit or implicit stratification were used
- All classrooms selected within school
- Classes were used as variance estimation strata and half classes were used to build jackknife replicates
- Did not participate in the Field Test

Allocation of School Sample in Macao SAR

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Macao SAR	57	0	57	0	0	0	0
Total	**57**	**0**	**57**	**0**	**0**	**0**	**0**

Malta

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, and foreign instructional language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (state, church, independent)
- No implicit stratification
- All classrooms were sampled
- All schools and all Grade 4 (Year 5) students were selected
- Classes were used as variance estimation strata and half classes were used to build jackknife replicates. All classrooms selected within schools.

Allocation of School Sample in Malta

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Church	25	0	25	0	0	0	0
Independent	8	0	8	0	0	0	2
State	62	0	62	0	0	0	0
Total	**95**	**0**	**95**	**0**	**0**	**0**	**2**

Morocco

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6)
- No within-school exclusions

Sample Design

- Explicit stratification by school type (private, public) and region (16)
- No implicit stratification
- The Field Test and Main Data Collection PIRLS samples were selected separately. The PIRLS Main Data Collection sample was selected by controlling for the overlap with the TIMSS 2015 samples and with the PIRLS Field Test sample using the Chowdhury approach.
- Oversampling of private schools and public within each region. All public schools were sampled in the region of Oued eddahab Lagouira. In these census strata, two classrooms were selected per school, and schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates. Sampled one classroom per school in other strata.
- PIRLS and PIRLS Literacy booklets were rotated within classes

Allocation of School Sample in Morocco

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Private - Grand Casablanca	12	0	12	0	0	0	0
Private - All Other Regions	28	0	28	0	0	0	0
Public - Chaouia Ouardigha	20	0	20	0	0	0	0
Public - Doukkala Abda	20	0	20	0	0	0	0
Public - Fes Boulmane	20	0	20	0	0	0	0
Public - Gharb Chrarda Beni Hssein	20	0	20	0	0	0	0
Public - Goulmim Smara	20	0	20	0	0	0	0
Public - Grand Casablanca	20	0	20	0	0	0	0
Public - Laayoune Boujdour Sakia Hamra	20	0	20	0	0	0	0
Public - Marrakech Tansift Haouz	20	0	20	0	0	0	0
Public - Meknes Tafilalt	20	0	20	0	0	0	0
Public - Oued eddahab Lagouira	21	1	20	0	0	0	0
Public - Rabat Salé Zemmour Zaer	20	0	20	0	0	0	0
Public - Région Est	20	0	20	0	0	0	0
Public - Souss Massa Draa	20	0	20	0	0	0	0
Public - Tadla Azilal	20	0	20	0	0	0	0
Public - Tanger Tetouan	20	0	20	0	0	0	0
Public - Taza Hoceima Taounate	20	0	20	0	0	0	0
Total	**361**	**1**	**360**	**0**	**0**	**0**	**0**

Netherlands

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6) and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities and non-native language speakers

Sample Design

- Explicit stratification by combinations of TIMSS and PIRLS socioeconomic status (5) and urbanization (5)
- No implicit stratification
- All classrooms were sampled
- PIRLS 2016 samples and TIMSS 2015 samples were selected simultaneously to avoid overlap

Allocation of School Sample in Netherlands

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
TIMSS & PIRLS High Mean SES - Very High Population Density	8	0	3	3	0	2	0
TIMSS & PIRLS High Mean SES - High Population Density	14	0	9	4	1	0	0
TIMSS & PIRLS High Mean SES - Moderate Population Density	16	0	11	3	0	2	0
TIMSS & PIRLS High Mean SES - Low Population Density	16	0	13	1	0	2	0
TIMSS & PIRLS High Mean SES - Very Low Population Density	16	0	14	1	1	0	0
TIMSS High & PIRLS Medium Mean SES - High to Very High Density	10	0	5	1	1	3	0
TIMSS High & PIRLS Medium Mean SES - Low to Moderate Density	14	0	9	1	3	1	0
TIMSS & PIRLS Medium Mean SES - High to Very High Density	10	0	7	3	0	0	0
TIMSS & PIRLS Medium Mean SES - Low to Moderate Density	12	1	11	0	0	0	0
TIMSS Medium & PIRLS Low Mean SES - High to Very High Density	14	1	10	2	0	1	0
TIMSS Medium & PIRLS Low Mean SES - Low to Moderate Density	10	0	8	1	0	1	0
TIMSS & PIRLS Low Mean SES	10	0	1	2	3	4	0
Total	**150**	**2**	**101**	**22**	**9**	**16**	**0**

New Zealand

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, Westmount closed Brethren campus, and correspondence school
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (4), socioeconomic status level (4), and urbanization (2)
- No implicit stratification
- Sampled two classrooms per school
- The PIRLS school samples were selected by controlling for the overlap with the TIMSS 2015 Grade 4 and Grade 8 samples using the Chowdhury approach

Allocation of School Sample in New Zealand

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Maori-Medium	10	0	4	1	1	4	0
English-Medium - High Immersion	10	0	8	1	0	1	0
Bilingual schools	8	0	5	0	1	2	0
English-Medium (other) - Independent	8	0	7	0	0	1	0
English-Medium (other) - Low SES	24	0	19	3	1	1	0
English-Medium (other) - Moderately low SES - Major urban centers	24	0	20	3	1	0	0
English-Medium (other) - Moderately low SES - Smaller centers	14	0	12	1	0	1	0
English-Medium (other) - Moderately high SES - Major urban centers	33	0	31	2	0	0	0
English-Medium (other) - Moderately high SES - Smaller centers	16	0	15	1	0	0	0
English-Medium (other) - High SES - Major urban centers	43	0	39	3	1	0	0
English-Medium (other) - High SES - Smaller centers	8	0	7	1	0	0	0
Total	**198**	**0**	**167**	**16**	**5**	**10**	**0**

Northern Ireland

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6) and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (5) and deprivation (5)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 58)
- PIRLS 2016 sample and TIMSS 2015 samples were drawn simultaneously to avoid overlap

Allocation of School Sample in Northern Ireland

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Belfast - Lower Deprivation Level	10	0	9	1	0	0	0
Belfast - Highest Deprivation Level	12	0	10	0	0	2	0
Western - Lower Deprivation Level	10	0	10	0	0	0	0
Western - Moderate to High Deprivation Level	10	0	5	1	0	4	0
Western - Highest Deprivation Level	8	0	6	0	0	2	0
North Eastern - Lowest Deprivation Level	8	0	6	0	0	2	0
North Eastern - Low to Moderate Deprivation Level	12	0	11	1	0	0	0
North Eastern - Higher Deprivation Level	14	0	12	1	0	1	0
South Eastern - Lowest Deprivation Level	12	1	9	0	0	2	0
South Eastern - Low to Moderate Deprivation Level	8	0	7	0	0	1	0
South Eastern - Higher Deprivation Level	14	0	13	0	0	1	0
Southern - Lower Deprivation Level	12	0	10	0	0	2	0
Southern - Moderate Deprivation Level	12	0	11	0	0	1	0
Southern - Higher Deprivation Level	12	0	11	0	0	1	0
Total	**154**	**1**	**130**	**4**	**0**	**19**	**0**

Norway (5)

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and instructional language other than Bokmal and Nynorsk
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by "Grade 5"/"Grade 4 and Grade 5" schools and language within "Grade 4 and Grade 5" stratum (Bokmål, Nynorsk)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 45)
- The PIRLS school samples were selected by controlling for the overlap with the TIMSS 2015 sample using the Chowdhury approach

Allocation of School Sample in Norway (5) - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Grade 5	7	1	6	0	0	0	0
Grade 4 & Grade 5 - Bokmål	126	0	119	5	0	2	0
Grade 4 & Grade 5 - Nynorsk	20	0	20	0	0	0	0
Total	**153**	**1**	**145**	**5**	**0**	**2**	**0**

Allocation of School Sample in Norway (5) - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Grade 5	7	1	6	0	0	0	0
Grade 4 & Grade 5 - Bokmål	126	0	114	4	0	8	0
Grade 4 & Grade 5 - Nynorsk	20	0	18	0	0	2	0
Total	**153**	**1**	**138**	**4**	**0**	**10**	**0**

Oman

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4) and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (government, private, international) and governorate (11) within government schools
- No implicit stratification
- In census strata and schools selected with certainty, schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates. Two classrooms selected within these schools. Sampled one classroom per school in other schools.

Allocation of School Sample in Oman

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Muscat Governorate	28	0	28	0	0	0	0
Ash Sharqiyah North Governorate	26	1	25	0	0	0	0
Ash Sharqiyah South Governorate	26	0	26	0	0	0	0
Ad Dakhliyah Governorate	26	0	26	0	0	0	0
Adh Dhahirah Governorate	26	0	26	0	0	0	0
Al Batinah North Governorate	30	0	30	0	0	0	0
Al Batinah South Governorate	26	0	26	0	0	0	0
Al Buraimi Governorate	15	0	15	0	0	0	0
Musandam Governorate	7	0	7	0	0	0	0
Al Wusta Governorate	20	0	20	0	0	0	0
Dhofar Governorate	26	0	26	0	0	0	0
Private Schools	26	0	24	1	0	1	0
International Schools	26	0	26	0	0	0	0
Total	**308**	**1**	**305**	**1**	**0**	**1**	**0**

Poland

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and instructional language other than Polish
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by urbanization (4) and school performance level (5)
- No implicit stratification
- Sampled two classrooms per school

Allocation of School Sample in Poland

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Village - Low Performance	14	0	14	0	0	0	0
Village - Medium-Low Performance	10	0	10	0	0	0	0
Village - Medium Performance	10	0	10	0	0	0	0
Village - Medium-High Performance	10	0	10	0	0	0	0
Village - High Performance	12	0	12	0	0	0	0
Town (Up to 20 thousand inhabitants) - Medium-Low Performance	8	0	8	0	0	0	0
Town (Up to 20 thousand inhabitants) - Medium-High Performance	10	1	8	1	0	0	0
City (20 to 100 thousand inhabitants) - Low Performance	10	0	9	1	0	0	0
City (20 to 100 thousand inhabitants) - Medium-Low Performance	8	0	7	0	0	1	0
City (20 to 100 thousand inhabitants) - Medium-High Performance	8	0	8	0	0	0	0
City (20 to 100 thousand inhabitants) - High Performance	10	0	9	1	0	0	0
City (Above 100 thousand inhabitants) - Low Performance	10	0	10	0	0	0	0
City (Above 100 thousand inhabitants) - Medium-Low Performance	10	0	9	1	0	0	0

Allocation of School Sample in Poland (Continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
City (Above 100 thousand inhabitants) - Medium-High Performance	10	0	9	1	0	0	0
City (Above 100 thousand inhabitants) - High Performance	10	0	8	2	0	0	0
Total	**150**	**1**	**141**	**7**	**0**	**1**	**0**

Portugal

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6), special needs schools, and minority language schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private) and region (7) within public schools
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 149)
- The PIRLS samples were selected by controlling for the overlap with the TIMSS 2015 sample using the Chowdhury approach
- Probability proportional to (school) size systematic sampling was used in the 3 largest explicit strata, and systematic sampling selection with equal probabilities was used in all other strata

Allocation of School Sample in Portugal - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Private - Lisboa	8	0	7	1	0	0	0
Private - All Other Regions	11	0	10	1	0	0	1
Public - Alentejo	30	0	27	2	0	1	0
Public - Algarve	8	0	8	0	0	0	0
Public - Centro	48	0	48	0	0	0	0
Public - Lisboa	36	0	35	1	0	0	0
Public - Norte	64	0	61	1	0	2	0
Public - R. A. Açores	8	0	7	1	0	0	0
Public - R. A. Madeira	8	0	8	0	0	0	0
Total	**221**	**0**	**211**	**7**	**0**	**3**	**1**

Allocation of School Sample in Portugal - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Private - Lisboa	8	0	7	1	0	0	0
Private - All Other Regions	11	0	10	1	0	0	1
Public - Alentejo	30	0	27	2	0	1	0
Public - Algarve	8	0	8	0	0	0	0
Public - Centro	48	0	48	0	0	0	0
Public - Lisboa	36	0	35	1	0	0	0
Public - Norte	64	0	61	1	0	2	0
Public - R. A. Açores	8	0	7	1	0	0	0
Public - R. A. Madeira	8	0	8	0	0	0	0
Total	**221**	**0**	**211**	**7**	**0**	**3**	**1**

Qatar

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, and instructional language other than English and Arabic
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (independent, community, private) and gender (boys, girls) within independent schools
- No implicit stratification
- Sampled two classrooms per school
- Census of schools
- Schools or classrooms or half classrooms were used to build jackknife replicates for variance estimation

Allocation of School Sample in Qatar

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Independent - Boys	46	0	46	0	0	0	0
Independent - Girls	49	0	49	0	0	0	0
Community	17	0	17	0	0	0	0
Private	106	2	104	0	0	0	0
Total	**218**	**2**	**216**	**0**	**0**	**0**	**0**

Russian Federation

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4) and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (42)
- No implicit stratification
- Sampled two classrooms in large schools in Moscow City (measure of size > 270), one classroom otherwise
- An extra sampling stage (regions) was required prior to sampling schools. 28 of 69 regions were selected with probability proportional to the region size and 14 bigger regions were selected with certainty. While each certainty region itself is an explicit stratum, the other sampled regions make one large explicit stratum. In the large explicit stratum, a sample of schools was selected within each region.
- Within regions, schools were selected with probability proportional to (school) size systematic sampling. Schools were sorted (serpentine) by location (up to 7 levels) before being sorted by school size. The same region sample was used for both TIMSS and PIRLS.
- Within the certainty regions, schools were paired for variance calculation purposes. Otherwise, selected regions were paired for variance calculation purposes.

Allocation of School Sample in Russian Federation

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Sankt-Petersburg*	6	0	6	0	0	0	0
Mosco City*	14	0	14	0	0	0	0
Moscow Region*	10	0	10	0	0	0	0
Nizhni Novgorod Region*	4	0	4	0	0	0	0
Perm Territory*	4	0	4	0	0	0	0
Samara Region*	4	0	4	0	0	0	0
Republic of Tatarstan*	6	0	6	0	0	0	0
Republic of Bashkortostan*	8	0	8	0	0	0	0
Krasnodar Territory*	8	0	8	0	0	0	0
Rostov Region*	6	0	6	0	0	0	0
Chelyabinsk Region*	6	0	6	0	0	0	0
Sverdlovsk Region*	8	0	8	0	0	0	0
Krasnoyarsk Territory*	4	0	4	0	0	0	0
Republic of Dagestan*	6	0	6	0	0	0	0
Novgorod Region	4	0	4	0	0	0	0
Kaliningrad Region	4	0	4	0	0	0	0
Vologda Region	4	0	4	0	0	0	0
Voronezh Region	4	0	4	0	0	0	0
Vladimir Region	4	0	4	0	0	0	0
Tula Region	4	0	4	0	0	0	0
Bryansk Region	4	0	4	0	0	0	0
Ryazan Region	4	0	4	0	0	0	0
Kaluga Region	4	0	4	0	0	0	0
Republic of Marij El	4	0	4	0	0	0	0
Ulyanovsk Region	4	0	4	0	0	0	0
Chuvashi Republic	4	0	4	0	0	0	0
Orenburg Region	4	0	4	0	0	0	0
Saratov Region	4	0	4	0	0	0	0
Astrakhan Region	4	0	4	0	0	0	0
Kurgan Region	4	0	4	0	0	0	0
Khanty Mansijsk AD	4	0	4	0	0	0	0
Irkutsk Region	4	0	4	0	0	0	0

* Certainty Regions

Allocation of School Sample in Russian Federation (Continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Kemerovo Region	4	0	4	0	0	0	0
Novosibirsk Region	4	0	4	0	0	0	0
Altai Territory	4	0	4	0	0	0	0
Zabaikalsk Territory	4	0	4	0	0	0	0
Tomsk Region	4	0	4	0	0	0	0
Sakhalin Region	4	0	4	0	0	0	0
Republic of Sakha (Yakutia)	4	0	4	0	0	0	0
Primorski Territory	4	0	4	0	0	0	0
Stravropol Territory	4	0	4	0	0	0	0
Kabardino-Balkarian Republic	4	0	4	0	0	0	0
Total	**206**	**0**	**206**	**0**	**0**	**0**	**0**

Saudi Arabia

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 7) and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region and by gender (boys, girls) within larger regions
- No implicit stratification
- Sampled one classroom per school

Allocation of School Sample in Saudi Arabia

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Asir - Boys	8	0	8	0	0	0	0
Asir - Girls	8	0	8	0	0	0	0
Bahah	8	0	8	0	0	0	0
Eastern Region - Boys	12	0	10	1	1	0	0
Eastern Region - Girls	12	0	11	0	1	0	0
Hail	8	0	8	0	0	0	0
Jawf	8	0	8	0	0	0	0
Jizan	10	2	5	1	2	0	0
Madinah - Boys	8	0	8	0	0	0	0
Madinah - Girls	8	0	7	1	0	0	0
Makkah - Boys	20	0	18	2	0	0	0
Makkah - Girls	22	0	20	1	1	0	0
Najran	8	4	0	1	3	0	0
Northern Borders	8	0	8	0	0	0	0
Qassim	8	0	8	0	0	0	0
Riyadh	44	0	42	2	0	0	0
Tabuk	8	0	8	0	0	0	0
Total	**208**	**6**	**185**	**9**	**8**	**0**	**0**

Singapore

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools and private schools
- For PIRLS 2016, like in all previous cycles, Singapore took a census of all public schools with Grade 4 students. The sampling frame excluded private schools, which are largely foreign-system schools operating in Singapore and which serve predominantly international students. These foreign-system schools are fundamentally different from the public schools in many respects (e.g., language of instruction; school-calendar year).
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- No explicit stratification
- No implicit stratification
- Sampled two classrooms per school
- Census of all schools. Within schools, two half classrooms were sampled with probability proportional to the size of the classroom. Within selected classrooms, 19 students were randomly sampled.
- Schools were used as variance estimation strata and classes were used to build jackknife replicates

Allocation of School Sample in Singapore - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
None	177	0	177	0	0	0	0
Total	**177**	**0**	**177**	**0**	**0**	**0**	**0**

Allocation of School Sample in Singapore - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
None	177	0	177	0	0	0	0
Total	**177**	**0**	**177**	**0**	**0**	**0**	**0**

Slovak Republic

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, and taught in language other than Slovak and Hungarian
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by language (Slovak, Hungarian), socioeconomic status (less than 1% of students coming from lower socioeconomic status, less than 10% of students coming from lower socioeconomic status, 10% or more students from lower socioeconomic status), and region group (5) within Slovak language strata
- No implicit stratification
- Sampled two classrooms per school
- Field Test and Main Data Collection samples were selected separately. The PIRLS Main Data Collection sample was selected using the Chowdhury method to minimize overlap with the PIRLS Field Test sample.
- Systematic sampling selection with equal probabilities used for sampling in strata with large sampling fractions

Allocation of School Sample in Slovak Republic

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Slovak - Higher SES - Region 1	20	0	19	1	0	0	0
Slovak - Higher SES - Regions 2 & 3 & 5	20	0	18	2	0	0	0
Slovak - Higher SES - Region 4	20	0	19	1	0	0	0
Slovak - Higher SES - Regions 6 & 8	16	1	14	1	0	0	0
Slovak - Higher SES - Region 7	20	0	19	1	0	0	0
Slovak - Medium and Lower SES - Region 1	7	0	7	0	0	0	0
Slovak - Medium SES - Regions 2 & 3 & 5	26	0	23	3	0	0	0
Slovak - Medium SES - Region 4	8	0	8	0	0	0	0
Slovak - Medium SES - Regions 6 & 7 & 8	20	0	19	1	0	0	0
Slovak - Lower SES - Regions 2 & 3 & 5	8	0	8	0	0	0	0
Slovak - Lower SES - Region 4	8	0	8	0	0	0	0
Slovak - Lower SES - Regions 6 & 7 & 8	32	0	31	0	1	0	0
Hungarian - Higher and Medium SES	8	0	7	1	0	0	0
Hungarian - Lower SES	8	0	8	0	0	0	0
Total	**221**	**1**	**208**	**11**	**1**	**0**	**0**

Slovenia

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools and Waldorf schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type according to school structure (main school, dislocated unit) and region (Pomurska, Koroška, Osrednjeslovenska, other regions)
- No implicit stratification
- Sampled two classrooms per school

Allocation of School Sample in Slovenia - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Main - Pomurska	10	1	9	0	0	0	0
Main - Koroška	10	0	10	0	0	0	0
Main - Osrednjeslovenska	26	1	24	0	0	1	0
Main - Other Regions	70	0	68	0	0	2	0
Dislocated - Pomurska	8	0	8	0	0	0	0
Dislocated - Koroška	13	0	13	0	0	0	0
Dislocated - Osrednjeslovenska	13	0	11	0	0	2	0
Dislocated - Other Regions	22	0	17	0	0	5	0
Total	**172**	**2**	**160**	**0**	**0**	**10**	**0**

Allocation of School Sample in Slovenia - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Main - Pomurska	10	1	9	0	0	0	0
Main - Koroška	10	0	10	0	0	0	0
Main - Osrednjeslovenska	26	1	24	0	0	1	0
Main - Other Regions	70	0	68	0	0	2	0
Dislocated - Pomurska	8	0	8	0	0	0	0
Dislocated - Koroška	13	0	12	0	0	1	0
Dislocated - Osrednjeslovenska	13	0	11	0	0	2	0
Dislocated - Other Regions	22	0	17	0	0	5	0
Total	**172**	**2**	**159**	**0**	**0**	**11**	**0**

South Africa

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, very small schools (measure of size < 6), schools for which language of testing cannot be determined, and schools with less than 30 learners
- Within-school exclusions consisted of students with intellectual disabilities and non-native language speakers

Sample Design

- Explicit stratification by language (11) and province (9)
- No implicit stratification
- Sampled two classrooms or more in schools teaching in more than one language
- Class group option was used in schools teaching in more than one language

Allocation of School Sample in South Africa

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Afrikaans - Northern Cape	10	0	10	0	0	0	0
Afrikaans - All other provinces	14	0	12	1	0	1	0
English - EC, GT, KZ, LP provinces	22	0	19	2	0	1	0
English - All other provinces	16	1	15	0	0	0	0
IsiNdebele - All provinces	8	0	8	0	0	0	0
IsiXhosa - Eastern Cape	16	0	16	0	0	0	0
IsiXhosa - All other provinces	8	0	8	0	0	0	0
IsiZulu - KwaZulu-Natal	24	0	18	4	1	1	0
IsiZulu - All other provinces	8	0	7	0	0	1	0
Sepedi - All provinces	16	0	15	0	0	1	0
Sesotho - All provinces	16	0	16	0	0	0	0
Setswana - Northern Cape	8	0	8	0	0	0	0
Setswana - All other provinces	20	0	20	0	0	0	0
SiSwati - All provinces	22	0	21	0	0	1	0
Tshivenda - All provinces	22	0	22	0	0	0	0
Xitsonga - All provinces	18	0	17	0	0	1	0
Afrikaans & English - EC, GT, KZ, LP provinces	4	0	4	0	0	0	0
Afrikaans & English - Northern Cape	4	0	2	0	0	2	0
Afrikaans & English - All other provinces	10	0	9	1	0	0	0
Neither Afrikaans nor English - FS & NC provinces	4	0	4	0	0	0	0

Allocation of School Sample in South Africa (Continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Neither Afrikaans nor English - All other provinces	12	0	12	0	0	0	0
Afrikaans/English/ others - EC, GT, KZ, LP provinces	12	0	11	1	0	0	0
Afrikaans/English/ others - All other provinces	10	1	8	1	0	0	0
Total	**304**	**2**	**282**	**10**	**1**	**9**	**0**

Spain

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, and international schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (8), school type (public, private). Within Madrid, private schools were also stratified by category (government dependent, independent) and by bilingual status (bilingual, not bilingual) within the public and government dependent private schools
- No implicit stratification
- Sampled two classrooms in large schools of Andalusia (measure of size > 74) and one classroom otherwise
- Oversampling of schools in Andalusia, Asturias, Basque Country, Canary Islands, Castile and Leon, Catalonia, La Rioja, and Madrid

Allocation of School Sample in Spain

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Andalusia - Public	110	0	109	1	0	0	0
Andalusia - Private	40	0	39	0	1	0	0
Asturias - Public	30	0	30	0	0	0	0
Asturias - Private	20	0	20	0	0	0	0
Castile and Leon - Public	29	0	29	0	0	0	1
Castile and Leon - Private	20	0	20	0	0	0	0
Catalonia - Public	30	0	30	0	0	0	0
Catalonia - Private	20	0	19	1	0	0	0
La Rioja - Public	27	0	27	0	0	0	0
La Rioja - Private	23	0	23	0	0	0	0
Madrid - Public - Bilingual	40	0	40	0	0	0	0
Madrid - Public - Non Bilingual	40	0	40	0	0	0	0
Madrid - Private - Bilingual	40	0	40	0	0	0	0
Madrid - Private - Non Bilingual	40	0	40	0	0	0	0
Madrid - Independent Private - Non Bilingual	8	0	8	0	0	0	0
Basque Country - Public	30	0	30	0	0	0	0
Basque Country - Private	20	0	20	0	0	0	0
Other regions - Public	42	0	41	1	0	0	0
Other regions - Private	20	0	20	0	0	0	0
Total	**629**	**0**	**625**	**3**	**1**	**0**	**1**

Sweden

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, international schools, special program schools, and very small schools (measure of size < 5)
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by grade average (4)
- No implicit stratification
- Sampled two classrooms per school
- The PIRLS sample was selected by controlling for the overlap with the TIMSS Grade 4 and Grade 8 samples using the Chowdhury approach

Allocation of School Sample in Sweden - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Higher average score	14	0	14	0	0	0	0
Medium average score	14	0	14	0	0	0	0
Low average score	24	1	23	0	0	0	0
Missing score	106	3	102	1	0	0	0
Total	**158**	**4**	**153**	**1**	**0**	**0**	**0**

Allocation of School Sample in Sweden - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Higher average score	14	0	13	0	0	1	0
Medium average score	14	0	13	0	0	1	0
Low average score	24	1	22	0	0	1	0
Missing score	106	3	96	0	0	7	0
Total	**158**	**4**	**144**	**0**	**0**	**10**	**0**

Trinidad and Tobago

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5)
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by type of school (government-related, private) and region within government-related stratum (8). Government-related strata include government and denominational schools.
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 69)

Allocation of School Sample in Trinidad and Tobago

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Private	12	0	12	0	0	0	0
Government-related - Caroni	22	0	22	0	0	0	0
Government-related - North Eastern	10	0	10	0	0	0	0
Government-related - Port of Spain and surroundings	20	0	20	0	0	0	0
Government-related - South Eastern	12	0	12	0	0	0	0
Government-related - St George East	32	0	32	0	0	0	0
Government-related - St. Patrick	16	1	15	0	0	0	0
Government-related - Tobago	8	0	8	0	0	0	0
Government-related - Victoria	20	0	20	0	0	0	0
Total	**152**	**1**	**151**	**0**	**0**	**0**	**0**

United Arab Emirates

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools, measure of size < 13 for all Emirates except Dubai and Abu Dhabi and measure of size < 10 for Dubai, schools with an instructional language other than Arabic, English, or French for Dubai and with an instructional language other than English and Arabic for the other Emirates, geographically inaccessible schools in all Emirates except Dubai, and home schools in Emirates other than Abu Dhabi and Dubai
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by Emirates (7), school type (public, private) and language of instruction (Arabic, English)
- No implicit stratification
- Census of schools in Dubai, Umm Al Quwain, and Fujairah private schools. Also, all private English schools with curriculum not from the United Kingdom, United States, or Canada, in the regions Abu Dhabi and Al Ain were sampled. In census strata, classes or half classes were used to build jackknife replicates for variance estimation. Two classrooms selected within these schools. Some schools are paired together within an explicit stratum when there is only one class participating.

Allocation of School Sample in United Arab Emirates - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public - Arabic	28	1	27	0	0	0	0
Private - Arabic	9	1	8	0	0	0	0
Private - English	138	1	136	0	0	1	0
Private - French	3	0	3	0	0	0	0
Abu Dhabi - Public - Both - ADEC Schools	26	0	26	0	0	0	0
Abu Dhabi - Private - Arabic - Ministry of Education	14	0	14	0	0	0	0
Abu Dhabi - Private - English - UK/US/CAD	30	1	29	0	0	0	0
Abu Dhabi - Private - English - Others	18	1	17	0	0	0	0
Al Ain - Public - Both - ADEC Schools	22	0	22	0	0	0	0
Al Ain - Private - Arabic - Ministry of Education	9	0	9	0	0	0	0
Al Ain - Private - English - UK/US/CAD	12	0	12	0	0	0	0
Al Ain - Private - English - Others	10	0	10	0	0	0	0
Al Gharbia	12	0	12	0	0	0	0
Sharjah - Public -Arabic	12	0	11	0	0	1	0
Sharjah - Private - Arabic	12	0	10	1	0	1	0
Sharjah - Private - English	20	0	20	0	0	0	0
Ajman - Public - Arabic	12	0	10	0	0	2	0
Ajman - Private - Arabic	12	0	12	0	0	0	0
Ajman - Private - English	8	0	8	0	0	0	0
Umm Al Quwain - Public - Arabic	6	0	6	0	0	0	0
Umm Al Quwain - Private - Arabic	1	0	1	0	0	0	0

Allocation of School Sample in United Arab Emirates - PIRLS (Continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Umm Al Quwain - Private - English	4	0	4	0	0	0	0
Fujairah - Public - Arabic	18	0	18	0	0	0	0
Fujairah - Private - Arabic	5	0	4	0	0	1	0
Fujairah - Private - English	9	0	9	0	0	0	0
Ras Al Khaimah - Public - Arabic	16	1	15	0	0	0	0
Ras Al Khaimah - Private - Arabic	8	0	7	0	0	1	0
Ras Al Khaimah - Private - English	8	1	7	0	0	0	0
Total	**482**	**7**	**467**	**1**	**0**	**7**	**0**

Allocation of School Sample in United Arab Emirates - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public - Arabic	28	1	27	0	0	0	0
Private - Arabic	9	1	8	0	0	0	0
Private - English	138	1	136	0	0	1	0
Private - French	3	0	3	0	0	0	0
Abu Dhabi - Public - Both - ADEC Schools	26	0	26	0	0	0	0
Abu Dhabi - Private - Arabic - Ministry of Education	14	0	14	0	0	0	0
Abu Dhabi - Private - English - UK/US/ CAD	30	1	29	0	0	0	0
Abu Dhabi - Private - English - Others	18	1	17	0	0	0	0
Al Ain - Public - Both - ADEC Schools	22	0	22	0	0	0	0
Al Ain - Private - Arabic - Ministry of Education	9	0	9	0	0	0	0
Al Ain - Private - English - UK/US/ CAD	12	0	12	0	0	0	0
Al Ain - Private - English - Others	10	0	10	0	0	0	0
Al Gharbia	12	0	11	0	0	1	0
Sharjah - Public - Arabic	12	0	11	0	0	1	0
Sharjah - Private - Arabic	12	0	10	1	0	1	0
Sharjah - Private - English	20	0	20	0	0	0	0
Ajman - Public - Arabic	12	0	9	0	0	3	0
Ajman - Private - Arabic	12	0	12	0	0	0	0
Ajman - Private - English	8	0	8	0	0	0	0
Umm Al Quwain - Public - Arabic	6	0	6	0	0	0	0
Umm Al Quwain - Private - Arabic	1	0	1	0	0	0	0

Allocation of School Sample in United Arab Emirates - ePIRLS (Continued)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Umm Al Quwain - Private - English	4	0	3	0	0	1	0
Fujairah - Public - Arabic	18	0	18	0	0	0	0
Fujairah - Private - Arabic	5	0	4	0	0	1	0
Fujairah - Private - English	9	0	9	0	0	0	0
Ras Al Khaimah - Public - Arabic	16	1	15	0	0	0	0
Ras Al Khaimah - Private - Arabic	8	0	7	0	0	1	0
Ras Al Khaimah - Private - English	8	1	7	0	0	0	0
Total	**482**	**7**	**464**	**1**	**0**	**10**	**0**

United States

Coverage and Exclusions

- Coverage is 100 percent
- No school level exclusions
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by poverty level (high, low), school type (public, private), and census region (4)
- Implicit stratification by urbanization (city, suburb, town, rural) and ethnicity status (above 15% non-White students in a school, below 15% non-White students in a school)
- Sampled two classrooms in large schools
- High poverty level schools were oversampled

Allocation of School Sample in United States - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
High Poverty Level - Public - Northeast	8	0	7	1	0	0	0
High Poverty Level - Public - Midwest	9	0	6	1	0	2	0
High Poverty Level - Public - South	24	0	23	1	0	0	0
High Poverty Level - Public - West	9	0	6	0	0	3	0
Low Poverty Level - Private - Northeast	3	1	2	0	0	0	0
Low Poverty Level - Private - Midwest	3	0	2	1	0	0	0
Low Poverty Level - Private - South	4	0	4	0	0	0	0
Low Poverty Level - Private - West	2	1	0	0	0	1	0
Low Poverty Level - Public - Northeast	18	1	11	3	0	3	0
Low Poverty Level - Public - Midwest	25	0	15	5	3	2	0
Low Poverty Level - Public - South	41	1	35	3	2	0	0
Low Poverty Level - Public - West	30	0	20	6	1	3	0
Total	**176**	**4**	**131**	**21**	**6**	**14**	**0**

Allocation of School Sample in United States - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
High Poverty Level - Public - Northeast	8	0	7	0	0	1	0
High Poverty Level - Public - Midwest	9	0	6	1	0	2	0
High Poverty Level - Public - South	24	0	22	1	0	1	0
High Poverty Level - Public - West	9	0	6	0	0	3	0
Low Poverty Level - Private - Northeast	3	1	2	0	0	0	0
Low Poverty Level - Private - Midwest	3	0	2	1	0	0	0
Low Poverty Level - Private - South	4	0	4	0	0	0	0
Low Poverty Level - Private - West	2	1	0	0	0	1	0
Low Poverty Level - Public - Northeast	18	1	11	3	0	3	0
Low Poverty Level - Public - Midwest	25	0	15	5	2	3	0
Low Poverty Level - Public - South	41	1	35	3	2	0	0
Low Poverty Level - Public - West	30	0	18	6	1	5	0
Total	**176**	**4**	**128**	**20**	**5**	**19**	**0**

Characteristics of Benchmarking Participants

Buenos Aires, Argentina

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of distance learning schools and special education schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private) and socioeconomic status (low, medium, high)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 70)

Allocation of School Sample in Buenos Aires, Argentina

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
State - Low SES	32	0	25	7	0	0	0
State - Medium SES	31	0	30	1	0	0	0
State - High SES	15	0	10	5	0	0	0
Private - Low SES	18	0	16	2	0	0	0
Private - Medium SES	27	0	24	3	0	0	0
Private - High SES	27	0	26	1	0	0	0
Total	**150**	**0**	**131**	**19**	**0**	**0**	**0**

Ontario, Canada

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 6), special needs schools, and First Nations schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (private, Catholic, public) and language (English, French) within Catholic and public schools
- Implicit stratification by region (4) in public and Catholic explicit strata
- Sampled two classrooms in large schools (measure of size > 79)
- The school sample for PIRLS was selected by controlling for the overlap with the TIMSS Grade 4 using the Chowdhury approach

Allocation of School Sample in Ontario, Canada

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Private	8	0	0	1	0	7	0
English - Catholic	30	0	30	0	0	0	0
English - Public	80	2	77	1	0	0	0
French - Catholic & Public	80	0	79	0	0	1	0
Total	**198**	**2**	**186**	**2**	**0**	**8**	**0**

Quebec, Canada

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), special needs schools, international schools, non ministry schools, and special status schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private) and language (French, English)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 80)
- The school sample for PIRLS was selected by controlling for the overlap with the TIMSS Grade 4 using the Chowdhury approach

Allocation of School Sample in Quebec, Canada

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
English - Private	8	1	7	0	0	0	0
English - Public	42	0	39	0	0	3	0
French - Private	8	0	8	0	0	0	0
French - Public	118	1	35	25	13	44	0
Total	**176**	**2**	**89**	**25**	**13**	**47**	**0**

Denmark (3)

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, daycare and rehabilitation home schools as well as German, English, and Rudolf Steiner schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (2)
- No implicit stratification
- Sampled one classroom per school
- The same sample of schools for PIRLS Grade 4 was used for Grade 3

Allocation of School Sample in Denmark (3)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public	171	7	154	9	0	1	0
Private	27	0	16	6	1	4	0
Total	**198**	**7**	**170**	**15**	**1**	**5**	**0**

Norway (4)

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 5), special needs schools, instructional language other than Bokmal and Nynorsk, and school for adults
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by "Grade 4"/"Grade 4 and Grade 5" schools and language within "Grade 4 and Grade 5" (Bokmål, Nynorsk)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 45)
- The PIRLS school samples were selected by controlling for the overlap with the TIMSS 2015 sample using the Chowdhury approach

Allocation of School Sample in Norway (4)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Grade 4	9	0	7	2	0	0	0
Grade 4 and Grade 5 - Bokmål	126	0	120	5	0	1	0
Grade 4 and Grade 5 - Nynorsk	20	0	20	0	0	0	0
Total	**155**	**0**	**147**	**7**	**0**	**1**	**0**

Moscow City, Russian Federation

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4) and special needs schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- No explicit stratification
- No implicit stratification
- Sampled 2 classrooms in large schools (measure of size > 270)

Allocation of School Sample in Moscow City, Russian Federation

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Moscow City	150	0	150	0	0	0	0
Total	**150**	**0**	**150**	**0**	**0**	**0**	**0**

Eng/Afr/Zulu – RSA (5)

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools, very small schools (measure of size < 6), schools with less than 30 learners, and Afrikaans & IsiZulu & English schools
- Within-school exclusions consisted of non-native language speakers

Sample Design

- Explicit stratification by language (Afrikaans only, English only, IsiZulu only, Afrikaans and English schools, IsiZulu and English schools)
- No implicit stratification
- Sampled two classrooms in bilingual schools
- The PIRLS Grade 5 sample was selected by controlling for the overlap with the Grade 4 PIRLS Literacy sample using the Chowdhury approach
- Class group option was used in bilingual schools

Allocation of School Sample in Eng/Afr/Zulu – RSA (5)

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Afrikaans - No English, No IsiZulu	24	0	20	1	0	3	0
English - No Afrikaans, No IsiZulu	45	10	29	1	0	5	0
IsiZulu - No Afrikaans, No English	49	1	41	1	3	3	0
Afrikaans & English	25	1	19	2	0	3	0
IsiZulu & English	9	1	8	0	0	0	0
Total	**152**	**13**	**117**	**5**	**3**	**14**	**0**

Andalusia, Spain

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 4), special needs schools, and international schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private)
- No implicit stratification
- Sampled two classrooms in large schools (measure of size > 74)

Allocation of School Sample in Andalusia, Spain

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public	110	0	109	1	0	0	0
Private	40	0	39	0	1	0	0
Total	**150**	**0**	**148**	**1**	**1**	**0**	**0**

Madrid, Spain

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of special needs schools and international schools
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, government dependent private, independent private) and bilingual status (bilingual, non bilingual)
- No implicit stratification
- Sampled one classroom per school

Allocation of School Sample in Madrid, Spain

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public - Bilingual	40	0	40	0	0	0	0
Public - Non Bilingual	40	0	40	0	0	0	0
Private - Bilingual	40	0	40	0	0	0	0
Private - Non Bilingual	40	0	40	0	0	0	0
Independent Private - Non Bilingual	8	0	8	0	0	0	0
Total	**168**	**0**	**168**	**0**	**0**	**0**	**0**

Abu Dhabi, United Arab Emirates

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of remote schools, and schools with an instructional language other than Arabic or English
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by region (Abu Dhabi, Al Ain, Al Gharbia), school type (public, private), language (Arabic, English), and curriculum (4)
- No implicit stratification
- All Private English schools with curriculum not from United Kingdom, United States, or Canada, were sampled in the regions Abu Dhabi and Al Ain. Two classrooms selected within these schools whenever possible. In these census strata, classes or half classes were used to build jackknife replicates for variance estimation. Sampled one classroom per school in other strata.

Allocation of School Sample in Abu Dhabi, United Arab Emirates - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Abu Dhabi - Public - Both - ADEC schools	26	0	26	0	0	0	0
Abu Dhabi - Private - Arabic - Ministry of Education	14	0	14	0	0	0	0
Abu Dhabi - Private - English - UK/US/CAD	30	1	29	0	0	0	0
Abu Dhabi - Private - English - Others	18	1	17	0	0	0	0
Al Ain - Public - Both - ADEC schools	22	0	22	0	0	0	0
Al Ain - Private - Arabic - Ministry of Education	9	0	9	0	0	0	0
Al Ain - Private - English - UK/US/CAD	12	0	12	0	0	0	0
Al Ain - Private - English - Others	10	0	10	0	0	0	0
Al Gharbia	12	0	12	0	0	0	0
Total	**153**	**2**	**151**	**0**	**0**	**0**	**0**

Allocation of School Sample in Abu Dhabi, United Arab Emirates - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Abu Dhabi - Public - Both - ADEC schools	26	0	26	0	0	0	0
Abu Dhabi - Private - Arabic - Ministry of Education	14	0	14	0	0	0	0
Abu Dhabi - Private - English - UK/US/CAD	30	1	29	0	0	0	0
Abu Dhabi - Private - English - Others	18	1	17	0	0	0	0
Al Ain - Public - Both - ADEC schools	22	0	22	0	0	0	0
Al Ain - Private - Arabic - Ministry of Education	9	0	9	0	0	0	0
Al Ain - Private - English - UK/US/CAD	12	0	12	0	0	0	0
Al Ain - Private - English - Others	10	0	10	0	0	0	0
Al Gharbia	12	0	11	0	0	1	0
Total	**153**	**2**	**150**	**0**	**0**	**1**	**0**

Dubai, United Arab Emirates

Coverage and Exclusions

- Coverage is 100 percent
- School-level exclusions consisted of very small schools (measure of size < 10), and schools with an instructional language other than Arabic, English, or French
- Within-school exclusions consisted of students with intellectual disabilities, students with functional disabilities, and non-native language speakers

Sample Design

- Explicit stratification by school type (public, private) and language (Arabic, English, French)
- No implicit stratification
- Sampled two classrooms per school
- Census of all schools
- Schools or classes were used as variance estimation strata and classes or half classes were used to build jackknife replicates

Allocation of School Sample in Dubai, United Arab Emirates - PIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public - Arabic	28	1	27	0	0	0	0
Private - Arabic	9	1	8	0	0	0	0
Private - English	138	1	136	0	0	1	0
Private - French	3	0	3	0	0	0	0
Total	**178**	**3**	**174**	**0**	**0**	**1**	**0**

Allocation of School Sample in Dubai, United Arab Emirates - ePIRLS

Explicit Strata	Total Sampled Schools	Ineligible Schools	Participating Schools			Refusal Schools	Excluded Schools
			Original Schools	1st Replacements	2nd Replacements		
Public - Arabic	28	1	27	0	0	0	0
Private - Arabic	9	1	8	0	0	0	0
Private - English	138	1	136	0	0	1	0
Private - French	3	0	3	0	0	0	0
Total	**178**	**3**	**174**	**0**	**0**	**1**	**0**

PIRLS
2016

Methods and Procedures in PIRLS 2016

Data Collection Procedures

CHAPTER 6

Survey Operations Procedures in PIRLS 2016

Ieva Johansone

Overview

As data-based indicators of countries' student achievement profiles and learning contexts, PIRLS assessments are crucially dependent on the quality of the data collected by each participating country and benchmarking entity. Whereas the development of the assessments is an intensely collaborative process involving all of the partners in the enterprise, the process of administering the assessments and collecting the data is uniquely the responsibility of each individual country or benchmarking participant.

To ensure the consistency and uniformity of approach necessary for high-quality, internationally comparable data, all participants are expected to follow a set of standardized operations procedures. These procedures have been developed through a partnership involving the TIMSS & PIRLS International Study Center, IEA Amsterdam, IEA Hamburg, Statistics Canada, and National Research Coordinators (NRCs) from participating countries. With each new assessment cycle, the operations procedures are updated to enhance efficiency and accuracy and reduce burden, making use of developments in information technology to automate routine activities wherever possible. Additionally, with the ePIRLS extension being administered for the first time in 2016, developing operations and procedures for this innovative assessment of online reading and integrating the workflow into the existing PIRLS operations was a significant undertaking.

In each country or benchmarking entity, the National Research Coordinator was responsible for the implementation of PIRLS 2016. Internationally, National Research Coordinators provided the country's perspective in all international discussions, represented the country at international meetings, and were the responsible contact persons for all project activities. Locally, National Research Coordinators were responsible for implementing all the internationally agreed-upon procedures and facilitating all of the national decisions regarding PIRLS, including any adaptations for the national context.

The daily tasks of the National Research Coordinators varied over the course of the PIRLS 2016 cycle. In the initial phases, National Research Coordinators participated in the PIRLS 2016 assessment framework and assessment development process (see Chapter 1) and collaborated with Statistics Canada and IEA Hamburg to develop a plan to implement the PIRLS 2016 sampling design within the country or benchmarking entity (see Chapter 5).

Following the development of the draft reading passages, achievement items, and context questionnaires, all countries conducted a full-scale field test of all instruments and operational procedures in March through April 2015 in preparation for the PIRLS 2016 data collection, which took place in October through December 2015 in Southern Hemisphere countries, and in March through May 2016 in Northern Hemisphere countries. The field test allowed the National Research Coordinators and their staff to become acquainted with the operational activities, and the feedback they provided was used to improve the procedures for the data collection. As expected, the field test resulted in some enhancements to survey operations procedures, especially for ePIRLS, which was new for the 2016 assessment cycle, and contributed to ensuring the successful execution of PIRLS 2016.

As part of ongoing efforts to improve operations, the National Research Coordinators were asked to complete a Survey Activities Questionnaire (SAQ), which sought feedback on all aspects of their experience conducting PIRLS 2016. The feedback solicited in the SAQ included an evaluation of the quality of the assessment materials and the effectiveness of the operations procedures and documentation. The results of the PIRLS 2016 Survey Activities Questionnaire are presented in the final section of this chapter.

PIRLS 2016 Survey Operations Units, Manuals, and Software

To support the National Research Coordinators in conducting the PIRLS 2016 assessments, the TIMSS & PIRLS International Study Center provided step-by-step documentation of all operational activities. Organized into a series of units, the *PIRLS 2016 Survey Operations Procedures* were made available at critical junctures of the project to ensure that National Research Coordinators had all the tools and information necessary to discharge their responsibilities. ePIRLS specific supplements to the PIRLS units were provided when necessary. Also, the *Procedures* units were accompanied by a series of manuals for use by School Coordinators and Test Administrators that National Research Coordinators could translate and adapt to their local situations. The TIMSS & PIRLS International Study Center and IEA Hamburg also provided National Research Coordinators and their staff with intensive training in constructed response item scoring and data management.

Consistent with the goal of automating and streamlining procedures wherever possible, IEA Hamburg provided National Research Coordinators with a range of custom-built software products to support activities, including sampling and tracking classes and students, administering school, teacher, and home questionnaires, documenting scoring reliability, and creating and checking data files. IEA Hamburg was also responsible for ePIRLS Software development. The ePIRLS system was hosted on the IEA Hamburg server and consisted of a number of software modules enabling the translation and verification processes, assessment administration to students, monitoring of the ePIRLS data upload, and scoring of the ePIRLS constructed response items.

The *Survey Operations Procedures* units were crucial resources for the National Research Coordinators as the units described in detail the tasks the NRCs were responsible for conducting. In the event that some of these tasks were contracted out to other people or organizations, the units ensured that the NRCs had sufficient knowledge of these matters to supervise the activities of the people who helped conduct the assessment(s) in their countries.

The following units, manuals, and software systems were provided for administering PIRLS and ePIRLS 2016:

- *PIRLS 2016 Survey Operations Procedures Unit 1: Sampling Schools and Obtaining their Cooperation*
- *PIRLS 2016 Survey Operations Procedures Unit 2: Preparing for and Conducting the PIRLS 2016 Field Test*. Unit 2 consisted of the following sections: Preparing Achievement Booklets and Background Questionnaires with an ePIRLS supplement on preparing the ePIRLS assessment tasks, Sampling Classes and Field Test Administration, Scoring the Constructed Response Items with an ePIRLS supplement on scoring the ePIRLS constructed response items online, and Creating the Field Test Databases. Unit 2 was accompanied by field test versions of the School Coordinator Manual, "Preparing Computers for ePIRLS" instructions, Test Administrator Manuals for PIRLS and ePIRLS, and a National Quality Control Monitor Manual. Eight software systems/modules (WinW3S, ePIRLS System Check Program, ePIRLS Online Translation System, ePIRLS Software, ePIRLS Online Data Monitor, ePIRLS Online Scoring System, IEA DME, and IEA OSS—described below) were provided for the field test.
- *PIRLS 2016 Survey Operations Procedures Unit 3: Contacting Schools and Sampling Classes for the Data Collection*. Unit 3 was accompanied by the School Coordinator Manual and the Windows® Within-school Sampling Software (WinW3S) and its manual. The WinW3S software enabled PIRLS 2016 participants to randomly select classes in each sampled school and document in detail the class selection process. The software also was used to track school, teacher, student, and student-teacher linkage information; prepare the survey tracking forms (described later in this chapter); and assign assessment instruments to students, including printing labels for the assessment instruments.

- *PIRLS 2016 Survey Operations Procedures Unit 4: Preparing Achievement Booklets and Context Questionnaires*. Unit 4 was accompanied by the IEA Online SurveySystem (OSS) and its manual. The IEA Online SurveySystem supported the online administration of the school, teacher, and home (Learning to Read Survey) questionnaires.
- *ePIRLS Supplement to the PIRLS 2016 Survey Operations Procedures Unit 4: Preparing the ePIRLS Assessment Tasks*. This ePIRLS supplement was accompanied by the ePIRLS Online Translation System enabling National Research Coordinators to connect to the ePIRLS server at IEA Hamburg to translate the ePIRLS assessment tasks into their language(s) of instruction. The translated tasks were then available online for translation and layout verification by IEA Hamburg and the TIMSS & PIRLS International Study Center (see Chapter 7).
- *PIRLS 2016 Survey Operations Procedures Unit 5: Conducting the Data Collection*. Unit 5 was accompanied by the Test Administrator Manuals for PIRLS and ePIRLS, the National Quality Control Monitor Manual, and the International Quality Control Monitor Manual.
- "Preparing Computers for ePIRLS" instructions and the ePIRLS System Check Program. The instructions and software provided the necessary information and tools for countries to test computers for ePIRLS compatibility and prepare the ePIRLS compatible computers for ePIRLS administration.
- ePIRLS Software for administering the ePIRLS assessment to students. ePIRLS Software was provided for each participating country and benchmarking entity individually, containing each participant's national/translated version of the ePIRLS assessment tasks.
- *PIRLS 2016 Survey Operations Procedures Unit 6: Scoring the Constructed Response Items*. Unit 6 was accompanied by the PIRLS 2016 Scoring Guides, the IEA Coding Expert Software, the Trend Reliability Scoring Manual, and the Cross-country Reliability Scoring Manual. The IEA Coding Expert Software was used to facilitate the trend and cross-country reliability scoring tasks.
- *ePIRLS Supplement to the PIRLS 2016 Survey Operations Procedures Unit 6: Scoring the Constructed Response Items*. This ePIRLS supplement was provided with the ePIRLS Online Data Monitor and ePIRLS Online Scoring System software facilitating monitoring of the ePIRLS data upload to the IEA Hamburg ePIRLS server and scoring the ePIRLS constructed response items.
- *PIRLS 2016 Survey Operations Procedures Unit 7: Creating the Databases*. Unit 7 was accompanied by the IEA Data Management Expert (DME) software, its manual, and codebooks that specified information on the IEA DME data fields in each of the data files. The IEA DME software is used for data entry and data verification.

PIRLS 2016 Survey Tracking Forms

PIRLS uses a series of tracking forms to document class sampling procedures, assign assessment instruments, and track school, teacher, and student information, including the participation status of the respondents. The tracking forms also facilitate the data collection and data verification process. Four different tracking forms were used for PIRLS 2016:

- Class Listing Form: This form was completed for each sampled school, listing the eligible classes and providing details about the classes, such as the class stream (if applicable), the number of students, and the names of teachers.
- Student Listing Form: This form was completed for each class sampled, listing the names of the students, student birth dates, gender, and exclusion codes.
- Student Tracking Form: This form was created for each class assessed and was completed by the Test Administrators during test administration. Separate Student Tracking Forms were provided for PIRLS and ePIRLS. The Test Administrators used this form to verify the assignment of survey instruments to students and to indicate participation status, including the return status of the Learning to Read Surveys (home questionnaires).
- Teacher Tracking Form: This form was completed for each sampled school to indicate the completion of the teacher questionnaires.

Operations for Data Collection

The following sections describe the major operational activities coordinated by the National Research Coordinators:

- Contacting schools and sampling classes
- Overseeing translation and preparing assessment instruments
- Managing the PIRLS 2016 assessment administration
- Scoring of the constructed response items
- Creating the PIRLS 2016 data files

Two other major PIRLS 2016 operational activities are described in separate chapters of the *Methods and Procedures in PIRLS 2016* publication—sampling schools (Chapter 3) and translation and layout verification of the assessment instruments (Chapter 7).

Contacting Schools and Sampling Classes

Exhibit 6.1 illustrates the major steps of working with schools to sample classes and prepare for the PIRLS 2016 assessment administration. Once the school samples were drawn, National Research Coordinators were tasked with contacting schools and encouraging them to take part in the assessment(s). Depending on the national context, this could involve obtaining support from national or regional educational authorities. *Survey Operations Procedures Unit 1* outlines suggestions on ways to encourage schools to participate in the assessment.

Exhibit 6.1: Diagram of the Sampling Procedures and Preparations for the Assessment Administration Implemented by National Centers and Schools

NATIONAL CENTER	SCHOOLS
Contacting and Tracking Schools • Contact sampled schools • Get started in WinW3S (complete project information and import the school sample database provided by Statistics Canada, translate / adapt tracking forms) • Complete / adapt school information • Record school participation • Print Class Listing Forms and send them to School Coordinators for completion	
	List all fourth grade classes and their teachers on the Class Listing Form
Class Sampling and Tracking; Preparing Computers for ePIRLS Administration • Enter school and class information from Class Listing Forms into WinW3S • Sample classes • Enter teacher information from Class Listing Forms into WinW3S • Print Student Listing Forms and send them to School Coordinators for completion • If school computers are used for ePIRLS administration, send the "Preparing Computers for ePIRLS" instructions and the ePIRLS System Check Program to School Coordinators	
	List student information on the Student Listing Forms. If applicable, run the ePIRLS System Check Program on all available computers.
Student and Teacher Tracking; Preparing Instruments for Assessment Administration • Confirm with School Coordinators the method for delivering the ePIRLS Software to students • Enter student information from Student Listing Forms into WinW3S • Assign achievement booklets and ePIRLS tasks to students • Print Student Tracking Forms • Print Teacher Tracking Forms • Print assessment instrument labels • Send tracking forms and labeled assessment materials to schools	
	ASSESSMENT ADMINISTRATION

In cooperation with school principals, National Research Coordinators were responsible for identifying and training School Coordinators for all participating schools. A School Coordinator could be a teacher or guidance counselor in the school, or National Research Coordinators could appoint a member of the national center to fill this role. In some countries, a School Coordinator from the national center was responsible for several schools in an area. School Coordinators were provided with a School Coordinator Manual, describing their responsibilities. The School Coordinator Manual was prepared by the TIMSS & PIRLS International Study Center and translated/adapted by national center staff in each country.

The responsibilities of the School Coordinators included providing the national center with information on the school; coordinating the dates, times, and places for testing; identifying and training Test Administrators to administer the assessments; coordinating the completion of the tracking forms; distributing questionnaires; and when necessary obtaining parental permission. If school computers were used for ePIRLS administration, School Coordinators were provided with the “Preparing Computers for ePIRLS” instructions and the ePIRLS System Check Program in order to test the computers for ePIRLS compatibility and prepare the compatible computers for testing. School Coordinators also confirmed receipt of all assessment materials, oversaw the security of the assessment materials, and ensured the return of the assessment materials to the national center following assessment administration.

In addition, School Coordinators provided the national center with data on eligible classes in the schools. With this information, the national centers used WinW3S to sample classes within the schools. Because PIRLS samples intact classes, the School Coordinators checked that every student was listed in one and only one class. This was necessary to ensure that the sample of classes resulted in a representative sample of students, and every student at the target grade had a chance of being selected.

Overseeing Translation and Preparing Assessment Instruments

National Research Coordinators also were responsible for preparing the assessment instruments (achievement booklets, ePIRLS tasks, and context questionnaires) for their countries—a process that included overseeing the translation of the assessment instruments. The overarching goal of assessment instrument preparation is to create internationally comparable instruments that are appropriately adapted for the national context of each participating country.

Each student was assigned one achievement booklet. There are 16 PIRLS achievement booklets and 16 PIRLS Literacy achievement booklets. Each booklet contains two assessment blocks, each including a passage with a set of items. Even though each assessment block appeared in more than one booklet, from an operational perspective, each block needed to be translated only once. Countries used Adobe® InDesign® software to link the translated and adapted assessment blocks to the appropriate booklets. Automating this process through Adobe® InDesign® decreased the chances of human error in the production process.

Students participating in ePIRLS were assigned two of five ePIRLS assessment tasks. ePIRLS translations and/or adaptations were applied through the ePIRLS Online Translation System and then distributed and delivered to students via the ePIRLS Software.

As described in Chapter 1, ten new assessment blocks were developed for PIRLS and PIRLS Literacy 2016, with the new blocks replacing the ones released at the end of the previous assessment cycle. Also, five assessment tasks were developed for the new ePIRLS 2016 assessment. The new assessment blocks (PIRLS passages and ePIRLS tasks) tasks were all tried out through the field test to investigate the psychometric characteristics of the achievement items. The best assessment blocks were chosen and some edits were applied for the main data collection. Similarly, the context questionnaires were evaluated following the field test to gauge the validity and reliability of the various questionnaire scales.

All participating countries and benchmarking entities translated and/or adapted the newly developed assessment blocks into the test administration language and did the same for the questionnaires. Countries that did not participate in PIRLS/prePIRLS 2011 or PIRLS 2006 had to translate and/or adapt the assessment blocks used in previous assessments (trend blocks) into their language(s) in preparation for the 2016 assessment administration. Countries that had participated in PIRLS/prePIRLS 2011 and/or PIRLS 2006 were required to use the same translations they used in those cycles.

For both the field test and main data collection, the participating countries received the international version (English) of the achievement booklets and context questionnaires with all the necessary instrument production files, including fonts and graphics files. For ePIRLS, this was done via the ePIRLS Online Translation System. Instructions on how to use the materials to produce high-quality, standardized instruments, were included in the corresponding *Survey Operations Procedures* Unit.

Once translated and/or adapted, first for the field test and then again for the main data collection, the passage/tasks, items, and context questionnaires were submitted to IEA Amsterdam for translation verification. IEA worked with independent translators to evaluate each country's translations and, when deemed necessary, suggested changes to the text.

After the translation verification, National Research Coordinators applied the necessary changes, and copies of the instruments were submitted to the TIMSS & PIRLS International Study Center for layout verification and to review national adaptations. This review checked that each booklet, ePIRLS assessment task, and questionnaire conformed to the international format and that any adaptations made to the instruments did not unduly influence their international comparability.

Documenting National Adaptations

While preparing national assessment instruments, countries sometimes by necessity made adaptations to the international versions. All national adaptations to the international assessment instruments, other than direct translation, were documented. For the achievement booklets and context questionnaires, the National Adaptations Forms (NAFs) were used to capture this documentation. For ePIRLS, national adaptations were documented via the ePIRLS Online Translation System.

During the translation verification and layout review, the verifiers checked whether the national adaptations were likely to influence the ability to produce internationally comparable data for the items involved. Any questions raised were directed to the National Research Coordinator for consideration.

The documentation was completed and reviewed at various stages of preparing national assessment instruments. Version I of the forms and online documentation was completed during the internal translation and review process and sent along with the rest of the materials for international translation verification. After translation verification, the documentation (Version II) was updated in response to the translation verifier's comments, reflecting any changes resulting from the verification, and sent along with the national assessment instruments for layout verification. Following layout verification, the national instruments and documentation were finalized (Version III) and submitted to IEA and the TIMSS & PIRLS International Study Center.

Managing the Administration of the PIRLS 2016 Assessments

Printing, preparing, and distributing assessment materials to the participating schools required careful organization and planning on the part of the National Research Coordinators. The assessment materials were packaged and sent to the School Coordinators prior to testing, giving ample time for the School Coordinators to confirm the receipt and correctness of the materials. The School Questionnaire and Teacher Questionnaires were then distributed, and the other instruments were kept in a secure room until the testing date.

Each sampled class was assigned a Test Administrator(s) who followed procedures described in the PIRLS and/or ePIRLS Test Administrator Manual to administer the assessment and student questionnaire. Test Administrators were in most cases chosen and trained by School Coordinators, and in some cases, the School Coordinator doubled as the Test Administrator.

Test Administrators were responsible for distributing materials to the appropriate students, reading the instructions provided in the Test Administrator Manual to the students, and timing the sessions. WinW3S systematically assigned achievement booklets and ePIRLS assessment tasks and produced labels to facilitate the distribution of the assessment, and Test Administrators used the Student Tracking Form(s) and these labels to distribute the assessment instruments to the correct students and to document student participation. When a class had a participation rate below 90 percent, it was the School Coordinator's responsibility to hold a makeup session for the absent students before returning all of the testing materials to the national center. Using the Test Administration Form, the Test Administrators documented the timing of the testing sessions and also solicited information about anything out of the ordinary that took place during assessment administration.

The PIRLS achievement booklets consisted of two sections with each containing one assessment block, and ePIRLS consisted of two parts with each containing one assessment task. To complete each part of the test, students were allowed 40 minutes, and the time was strictly enforced by Test Administrators. ePIRLS Software also automatically logged students out of the system once the 40 minutes had expired. There was a required break between the two parts of assessment administration. The break was not to exceed 30 minutes. Students who completed part 1 or part 2 of the assessment before the allotted time were not allowed to leave the testing room and were asked to review their answers or read quietly. Some Test Administrators provided activity sheets for these students.

Following the administration of the PIRLS assessment, students were provided 30 minutes to complete the student questionnaire with extra time provided to students who needed it. During administration of the student questionnaire, Test Administrators were permitted to read the questionnaire items aloud together with the students. Following the administration of the ePIRLS assessment, students also took a short computer-based questionnaire about their experiences and attitudes toward using a computer.

PIRLS, including the student questionnaire, was always administered before ePIRLS. ePIRLS was mostly administered via individual USB sticks on individual ePIRLS compatible computers. Sometimes, the server method was used via a Local Area Network (LAN), which entailed a single ePIRLS compatible computer being used as a local server and students using individual devices connected to the server computer. For ePIRLS, the Test Administrators and School Coordinators submitted the ePIRLS data after each testing session. Due to computer shortages, sometimes multiple ePIRLS testing sessions were needed for each class.

Linking Students to their Teachers and Classes

Exhibit 6.2 illustrates the hierarchical identification system codes that are used to link the data among schools, classes, students, and teachers. The school, class, and student IDs are strictly hierarchical, with classes nested within schools and students nested within classes.

Exhibit 6.2: Hierarchical Identification System Codes Used to Link Schools, Classes, Students, and Teachers

Participant	ID Components	ID Structure	Numeric Example
School	School	CCCC	0001
Class	School + Class within the school	CCCCKK	000101 000102
Student	School + Class within the school + Student within the class	CCCCKKSS	00010101 00010201
Teacher	School + Teacher within the school + Linkage number to the sampled class	CCCCTTLL	00010101 00010201

Each teacher is assigned a teacher identification number consisting of the four-digit school number followed by a two-digit teacher number. Since the same teacher could be teaching more than one class within a school, it is necessary to have a unique identification number for each teacher linked to a class. This is achieved by adding a two-digit link number to the six digits of the teacher identification number to create a unique eight-digit identification number.

Online Administration of the School, Teacher, and Home Questionnaires

Countries could choose to administer the school, teacher, and home questionnaires online. The benefits of administering the questionnaires online included saving money and time in printing, and improving the efficiency of questionnaire distribution, data entry, and data cleaning.

For the online administration of the questionnaires, IEA Hamburg provided its IEA Online SurveySystem Software that incorporates design, presentation, and monitoring components.

The design component, known as the Designer, supports the preparation of the online surveys, data management, and data output to IEA Hamburg. Through the IEA Online SurveySystem Designer component, national centers could tailor the online questionnaires to their national language. To facilitate translation and adaptation, the Designer concurrently stored the original English question text and the translations and/or national adaptations. It also stored the variable names and data validation rules. If a national center decided not to administer a particular international question or option, it could be disabled in the Designer and would not be administered during the online questionnaire administration. The Designer also included an integrated preview function to allow for a visual side-by-side comparison of the paper/PDF and online versions of the questionnaires, facilitating the layout verification process.

For the online presentation, the Web Component presents the questionnaires to the respondents. The navigation capabilities of the Web Component are designed to allow respondents to pick and choose their order of response. Buttons marked "next" and "previous" facilitated navigation between adjacent pages, so users could browse through the questionnaire in the same way that they flip through the pages of the paper questionnaire. A hyperlinked interactive "table of contents" allowed the respondents to fluidly navigate to specific questions. Overall, these two functions permitted the respondents to answer questions in the order of their choosing, and skip questions just as they could do if they were answering the paper questionnaire. Also, the online questionnaires could be accessed through any standard Internet browser on all standard operating systems without the user needing any additional software.

Finally, the Web-based Monitor component allows for monitoring the survey responses in real time. Many national centers made extensive use of the Web-based Monitor to follow-up with non-respondents.

IEA Hamburg followed a stringent set of procedures to safeguard the confidentiality of the respondents and maintain the integrity of the data. Each respondent received a statement of confidentiality, and information on how to access the online questionnaire. For most countries, the online questionnaire administration was hosted on the IEA Hamburg customized high-performance server. This server allowed for the 24-hour availability of the questionnaires during the data collection period, and it also ensured backup and recovery provisions for the data.

Scoring the Constructed Response Items

Constructed response items represent a substantial portion of the PIRLS assessments, and because reliable and valid scoring of these items is critical to the assessment results, the TIMSS & PIRLS International Study Center provided explicit scoring guides and extensive training in their use. Also, the *Survey Operations Procedures* units specified a procedure for efficiently organizing and implementing the scoring activity. Scoring of the ePIRLS constructed response items was done online via the ePIRLS Online Scoring System, which incorporated the IEA standards and reliability procedures.

International scoring training sessions (one for the field test and two for the main data collection—one for Southern Hemisphere countries and another for Northern Hemisphere countries) were conducted where all National Research Coordinators (or country representatives appointed by the National Research Coordinators) were trained to score each of the constructed response items. At these training sessions, the scoring guide for each item was reviewed and applied to a sample set of example student responses that had already been scored. These example papers were chosen to represent a range of response types and to demonstrate the guides as clearly as possible. Following the example papers, the training participants applied the scoring guides to a different set of student responses that had not yet been scored. The scores to these practice papers were then shared with the group and any discrepancies were discussed.

Following the international scoring training, national centers trained their scoring staff on how to apply the scoring guides for the constructed response items. National Research Coordinators were encouraged to create additional example papers and practice papers from student responses collected in their country.

Documenting Scoring Reliability

Because reliable scoring of the constructed response items is essential for high quality data, it is important to document the reliability of the scoring process. A high degree of scorer agreement is evidence that scorers have applied the scoring guides in the same way. The procedure for scoring the PIRLS constructed response items provided for documenting scoring reliability within each country (within-country reliability scoring), over time (trend reliability scoring), and across countries (cross-country reliability scoring).

The method for establishing the reliability of the scoring within each country was for two independent scorers to score a random sample of 200 responses for each constructed response item. The degree of agreement between the scores assigned by the two scorers is a measure of the reliability of the scoring process. In collecting the within-country reliability data, it was vital that the scorers independently scored the items assigned to them, and each scorer did not have prior knowledge of the scores assigned by the other scorer. The within-country reliability scoring was integrated within the main scoring procedure and ongoing throughout the scoring process. The within-country reliability scoring procedure was implemented in both PIRLS and ePIRLS.

The purpose of the trend reliability scoring was to measure the reliability of the scoring from one assessment cycle to the next (i.e., from PIRLS 2011 to PIRLS 2016). The trend reliability scoring required scorers of PIRLS 2016 to score student responses collected in 2011. The scores from 2016 were then compared with the scores awarded in 2011. Trend reliability scoring was conducted using the IEA Coding Expert Software provided by IEA Hamburg.

Student responses included in the trend reliability scoring (150–200 responses per item) were actual student responses to 22 items from four of the PIRLS trend assessment blocks and/or 24 items from three of the PIRLS Literacy trend assessment blocks collected during the PIRLS/prePIRLS 2011 assessment administration in each country and benchmarking entity. These responses were scanned and provided for each participating country and benchmarking entity along with the IEA Coding Expert Software. All scorers who scored the trend assessment blocks in 2016 were required to participate in the trend reliability scoring. If all scorers were trained to score all trend items, the software divided the student responses equally among the scorers. If scorers were trained to score specific assessment blocks, National Research Coordinators were able to specify within the software which scorers would score particular blocks, and the software allocated the student responses accordingly. Similar to the within-country reliability scoring, the trend reliability scoring had to be integrated within the main scoring procedure.

Finally, cross-country reliability scoring gave an indication about how consistently the scoring guides were applied from one country to the next. The cross-country reliability scoring also was conducted using IEA Coding Expert Software. Student responses included in the cross-country reliability scoring (200 responses per item) were student responses to 22 items from four of the

PIRLS assessment blocks (the same passages and items were used for the trend scoring reliability study) that were collected from the English-speaking countries during the PIRLS 2011 assessment administration. All scorers who could score student responses written in English were required to participate in the cross-country reliability scoring, and the student responses were equally divided among the participating scorers in each country. The scoring exercise was completed immediately after all other scoring activities.

Creating the PIRLS 2016 Databases

The data entry process took place from March to May 2015 for the field test, from December 2015 to March 2016 following data collection in the Southern Hemisphere, and June to September 2016 following data collection in the Northern Hemisphere. The procedure for creating the PIRLS 2016 databases included entering sampling and assessment administration information into the WinW3S database and adding responses from the context questionnaires and achievement booklets using the IEA Data Management Expert (DME) software. IEA Hamburg provided the DME software to accommodate keyboard data entry from the paper instruments. The DME software also offers data and file management capabilities, a convenient checking and editing mechanism, interactive error detection, and quality control procedures.

Along with the DME software, IEA Hamburg provided international codebooks describing all variables and their characteristics, thus ensuring that the data files met the internationally defined rules and standards for data entry. The files within the DME database for entering the PIRLS 2016 data were based on these codebooks. However, the codebooks had to match exactly the national assessment instruments so that the answers of the respondents could be entered properly. Therefore, any adaptations to the international instruments also required adaptations to the international codebooks. The adapted national codebooks then were used to create the PIRLS 2016 data files in each country, with the responses to the context questionnaires, achievement booklets, and Reliability Scoring Sheets keyed into the DME database.

Quality control throughout the data entry process was essential to maintain accurate data. Therefore, National Research Coordinators were responsible for performing periodic reliability checks during data entry and for applying a series of data verification checks provided by both WinW3S and DME software prior to submitting the databases to IEA Hamburg. To ensure the reliability of the data entry process, the data entry staff was required to double enter at least 5 percent of each instrument type. An error rate of 1 percent or less was acceptable for the background files. An error rate of 0.1 percent or less was required for the student achievement files and the reliability scoring files. If the required agreement was not reached, retraining of the key punchers was required.

The ePIRLS assessment data were captured automatically by submitting them to the IEA Hamburg ePIRLS server immediately after the assessment administration. Countries were provided with the ePIRLS Online Data Monitor to monitor the data submission. The ePIRLS constructed response scoring took place directly in the online database and thus did not require any manual data entry. For the PIRLS 2016 teacher, school, and home questionnaires administered online through the Online SurveySystem (OSS) via the IEA Hamburg server, the data were directly accessible by IEA Hamburg and no further data entry was required.

Both WinW3S and DME offer a data verification module identifying a range of problems, such as inconsistencies of identification codes, inconsistencies between participation status information and achievement and/or background data availability, and out-of-range or otherwise invalid codes. The data quality control procedures also verify the integrity of the linkage between the students, teachers, and schools entered into the DME database and tracking of information for those specified in WinW3S. For data captured online (i.e., ePIRLS and context questionnaires administered online), it was possible to export data availability information and apply data verification to check for inconsistencies via the WinW3S and DME data verification modules.

When all data files had passed the quality control checks, they were submitted to IEA Hamburg, along with data documentation, for further checking and processing. For information on data processing at IEA Hamburg, please refer to Chapter 9 of this publication.

PIRLS 2016 Survey Activities Questionnaire

The Survey Activities Questionnaire was designed to elicit information about National Research Coordinators' experiences in preparing for and conducting the PIRLS 2016 data collection. The questionnaire was composed of six sections and focused on the following:

- Sampling schools and classes
- Preparing assessment instruments
- Administering the assessment(s)
- Implementing the National Quality Control Program
- Preparing for and scoring the constructed response items
- Creating the databases

All items in the Survey Activities Questionnaire included accompanying comment fields, in which NRC respondents were encouraged to explain their responses, provide additional information, and suggest improvements for the process.

The *PIRLS 2016 Survey Activities Questionnaire* was administered online via the IEA's Online SurveySystem and was completed by a total of 52 NRCs, with 15 NRCs also providing feedback on ePIRLS administration. The following sections summarize information gathered from the Survey Activities Questionnaire.

Sampling Schools and Classes

The first section of the Survey Activities Questionnaire asked National Research Coordinators about the *Survey Operations Procedures* Units for sampling both schools and classes within the sampled schools. As shown in Exhibit 6.3, all but one of the National Research Coordinators considered that *Survey Operations Procedures* Units 1 and 3 to be clear and sufficient. Two countries reported deviating from the basic PIRLS sampling design. Their reasons for these modifications to the sampling procedures included allowing for census participation, oversampling certain regions, and specific requirements to coordinate their PIRLS 2016 sample with the TIMSS 2015 sample. Statistics Canada, in cooperation with IEA Hamburg, selected the school samples for all countries and benchmarking participants.

Exhibit 6.3: Survey Activities Questionnaire, Section One—Sampling (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "PIRLS 2016 Survey Operations Procedures Unit 1 – Sampling Schools and Obtaining their Cooperation" clear and sufficient?	51	0	1
Were there any conditions or organizational constraints that necessitated deviations from the basic PIRLS sampling design described in the "Survey Operations Procedures Unit 1"?	2	49	1
Did you use the Within-school Sampling Software (WinW3S) to sample classes?	49	2	1
Did you experience any problems or inconveniences when using the WinW3S software?	16	32	4
Was the information provided in the "PIRLS 2016 Survey Operations Procedures Unit 3 – Contacting Schools and Sampling Classes for the Data Collection" clear and sufficient?	50	1	1
Did you follow the procedures outlined in "Survey Operations Procedures Unit 3" for working with the schools to sample classes (e.g., using the appropriate tracking forms in the proposed order to obtain information from School Coordinators)?	38	13	1

Two National Research Coordinators reported not using the Windows® Within-school Sampling Software (WinW3S) provided by IEA Hamburg to select classes within the sampled schools. One of them was for a benchmarking entity of a participating country, for which WinW3S was used centrally to sample classes within schools for the whole country. National Research Coordinators did report experiencing problems using the WinW3S Software. Among the issues reported were the slow speed of the software, the software not working on a shared network, issues importing information from Excel, problems with "right-to-left" languages, and issues coordinating PIRLS and ePIRLS participation status. National Research Coordinators also suggested that the software could be improved through the addition of an export to Excel function.

Thirteen National Research Coordinators applied some modifications to the procedures outlined in the *Survey Operations Procedures Unit 3*. For example, some National Research Coordinators did not use the Class Listing Forms because all classes at the target grade were tested or because a class level database was available at the ministry, and a number of countries did not use the Teacher Tracking Forms because there was only one teacher per class. All modifications were reviewed and approved by the TIMSS & PIRLS International Study Center.

Translating, Adapting, and Producing Assessment Instruments

The second section of the Survey Activities Questionnaire asked National Research Coordinators about translating, adapting, assembling, and printing the test materials, as well as issues related to checking the materials and securely storing them. Some ePIRLS specific questions were asked in this section that were related to using the ePIRLS Online Translation System, receiving ePIRLS Software, and preparing USBs in order to deliver ePIRLS to schools and students.

As reported in Exhibit 6.4, almost all National Research Coordinators found the instructions on preparing achievement booklets, context questionnaires, and ePIRLS assessment tasks clear and sufficient. However, ten countries reported experiencing some problems using the survey instrument production materials and/or the ePIRLS Online Translation System. These problems mostly included issues with fonts and special characters (e.g., for Cyrillic alphabet), difficulty fitting longer national text in the context questionnaires, and some problems with the layout style of tables. Among the problems reported about using the ePIRLS Online Translation System were inconsistencies between PDF storyboards and the translation system, some text not exporting properly to PDF, and the inability to hyphenate words. All of the identified problems were resolved either by specialists at the national center or with assistance from IEA Hamburg and the TIMSS & PIRLS International Study Center.

All National Research Coordinators, except one for a benchmarking entity of a participating country, reported applying corrections to their survey instruments as suggested by the external translation verifier or the layout verifier.

Exhibit 6.4: Survey Activities Questionnaire, Section Two—Translating, Adapting, and Producing Assessment Instruments (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "PIRLS 1016 Survey Operations Procedures Unit 4 – Preparing Achievement Booklets and Context Questionnaires" clear and sufficient?	49	2	1
Was the information provided in the "ePIRLS Supplement to the PIRLS 2016 Survey Operations Procedures Unit 4 – Preparing the ePIRLS Assessment Tasks" clear and sufficient?	15	0	0
Did you encounter any major problems using the assessment instrument production materials (e.g., instrument production files, fonts, support materials) provided by the TIMSS & PIRLS International Study Center?	4	47	1
Did you encounter any major problems with the ePIRLS Online Translation System?	6	9	0
After the translation verification, did you correct your translations/adaptations as suggested by the verifier in the majority of cases?			
PIRLS/PIRLS Literacy booklets	50	1	0 *(Not Answered)* 1 *(Not Applicable)*
Context Questionnaires	49	1	0 *(Not Answered)* 2 *(Not Applicable)*
ePIRLS assessment tasks	13	1	0 *(Not Answered)* 1 *(Not Applicable)*
After the layout verification, did you correct your assessment instruments as noted by the verifier in the majority of cases?			
PIRLS/PIRLS Literacy booklets	51	0	0 *(Not Answered)* 1 *(Not Applicable)*
Context Questionnaires	50	0	0 *(Not Answered)* 2 *(Not Applicable)*
ePIRLS assessment tasks	13	0	0 *(Not Answered)* 2 *(Not Applicable)*
Did you apply any quality control measures to check the achievement booklets and context questionnaires during the printing process (e.g., checking for missing pages, upside down pages, text too bright or too dark)?	49	3	0
Did you experience any problems receiving the final ePIRLS Software from IEA Hamburg and preparing the ePIRLS USB sticks?	2	13	0
Did you apply any quality control of the prepared ePIRLS USB sticks before sending them to the participating schools?	11	4	0
Did you take measures to protect the security of the assessment instruments during the translation, assembly, and printing process?	51	1	0
Did you detect any potential breaches in security of the assessment instruments?	0	51	0
Did you encounter any problems preparing the Online SurveySystem files for administering the school, teacher, and/or home (Early Learning Survey) questionnaires online?	1	22	0 *(Not Answered)* 29 *(Not Applicable)*

Nearly all of the countries conducted the recommended quality control checks during the process of printing the testing materials for PIRLS and preparing USBs for ePIRLS. The most common errors detected and fixed during the printing process were pages that were missing or in the wrong order. For ePIRLS, two countries reported issues with their initial ePIRLS Software, which was then corrected and new software provided.

One country reported that they experienced a problem with the IEA's Online SurveySystem (OSS). They reported that they could not print from the OSS web print preview.

Assessment Administration

The third section of the Survey Activities Questionnaire addressed the extent to which National Research Coordinators detected errors in the testing materials during packaging for shipment to schools. As shown in Exhibit 6.5, a small number of errors were found in the materials. Approximately half of such errors were corrected before distributing the materials to the respondents. Errors found after distribution usually were very minor, and either were fixed by School Coordinators or replacement materials were provided. The few cases where the errors could not be remedied were reported to the TIMSS & PIRLS International Study Center, where decisions were made about setting the problematic data to "Not Administered."

Exhibit 6.5: Survey Activities Questionnaire, Section Three—Assessment Administration (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "PIRLS 2016 Survey Operations Procedures Unit 5 – Conducting the Data Collection" clear and sufficient?	51	0	0
Were any errors detected in any of the following assessment materials after they were sent to schools?			
Achievement booklets	11	41	0 *(Not Answered)* 0 *(Not Applicable)*
Achievement booklet ID labels	6	45	0 *(Not Answered)* 1 *(Not Applicable)*
ePIRLS USB Sticks	0	15	0 *(Not Answered)* 0 *(Not Applicable)*
Student Questionnaires	7	44	0 *(Not Answered)* 1 *(Not Applicable)*
Student Questionnaire ID labels	5	45	0 *(Not Answered)* 2 *(Not Applicable)*
Learning to Read Surveys	3	45	0 *(Not Answered)* 4 *(Not Applicable)*
Learning to Read Survey ID labels	3	45	0 *(Not Answered)* 4 *(Not Applicable)*
Student Tracking Forms	3	48	0 *(Not Answered)* 1 *(Not Applicable)*
Teacher Questionnaires	0	50	0 *(Not Answered)* 2 *(Not Applicable)*

Exhibit 6.5: Survey Activities Questionnaire, Section Three—Assessment Administration (Numbers of NRC Responses) (Continued)

Question	Yes	No	Not Answered
Teacher Tracking Forms	0	48	0 *(Not Answered)* 4 *(Not Applicable)*
School Questionnaires	2	49	0 *(Not Answered)* 1 *(Not Applicable)*
School Coordinator Manuals	2	48	0 *(Not Answered)* 2 *(Not Applicable)*
Test Administrator Manuals	5	45	0 *(Not Answered)* 2 *(Not Applicable)*
If any errors were detected, did you correct the error(s) before the testing began?	17	18	0 *(Not Answered)* 17 *(Not Applicable)*
Does your country have a confidentiality policy that restricts putting student names on tracking forms and survey instrument covers?	13	39	0
Did you encounter any problems translating and/or adapting the School Coordinator Manual?	1	51	0
Did you encounter any problems translating and/or adapting the "Preparing Computers for ePIRLS" instructions?	0	15	0
Did you experience any software-specific problems when using the ePIRLS System Check Program to test computers for ePIRLS comparability?	2	13	0
Did you encounter any problems translating and/or adapting the Test Administrator Manual(s)?	2	50	0
Were School Coordinators appointed from within the participating schools?	44	8	0
Did you hold formal training session(s) for School Coordinators?	31	21	0
Were Test Administrators trained by School Coordinators within the participating schools?	30	22	0
Did Test Administrators document any problems or special circumstances that occurred frequently during the assessment administration (please refer to the completed Test Administration Forms)?	15	37	0
Did you require/suggest/provide an additional person to help the Test Administrator during the ePIRLS testing sessions?	14	1	0
Did you have a sufficient number of computers available for all/most schools to test all of the selected students (the whole class) at the same time?	8	7	0
Did you experience any software-specific problems when using the ePIRLS Software?	8	7	0
Did you use the individual computers/USB sticks or the server method to administer ePIRLS in your country?			
Individual computers/USB sticks	10	-	0
Server method	0	-	0
Both methods were used	5	-	0
Did you experience any software-specific problems when using the ePIRLS Online Data Monitor?	4	11	0
If you administered school, teacher, and/or home (Learning to Read Survey) questionnaires online, did any of the respondents in your country encounter any problems responding to the online questionnaires?	3	17	0 *(Not Answered)* 31 *(Not Applicable)*

Three National Research Coordinators reported difficulties translating the School Coordinator Manual and/or the Test Administrator Manual. Primarily, problems arose when the manual(s) had to be reorganized or adapted and the standardized procedures were modified (e.g., no Class Listing Forms or Teacher Tracking Forms were used).

Preparing computers for ePIRLS went smoothly—no country participating in ePIRLS reported problems translating and adapting the instructions provided and only two National Research Coordinators reported problems with the ePIRLS System Check Program. For these participants, changes had occurred on some of the computers in some schools after the initial system check, and the ePIRLS Software could not be run on these computers despite a successful initial ePIRLS compatibility test. One country reported problems with running the ePIRLS Software on Apple computers via the server method—their Apple computers could not open the ePIRLS Software.

In 44 countries, School Coordinators were appointed from within the participating schools and in the remaining countries, School Coordinators were from the national center or were contracted externally. In most countries, the National Research Coordinators organized training sessions for School Coordinators. In some, mostly larger countries, training was conducted either online or in a written form via extended manuals. In 30 countries, Test Administrators were trained by the School Coordinators within the participating schools.

Although the PIRLS administration when very well, Test Administrators occasionally reported difficulties. Among the problems documented by Test Administrators during assessment administration were the following: loud noises outside the classroom, many students asking questions, confusion about the PIRLS Reader and its booklet, too much time, not enough time, some technical problems with the ePIRLS administration, the student questionnaire being too long, confusion about the spare assessment materials, and student complaints that the test was too difficult.

In all but one country participating in ePIRLS, an additional person helped Test Administrators during the ePIRLS testing sessions. Half the ePIRLS countries had enough ePIRLS compatible computers to test all students in the participating classes at the same time, while the other half organized more than one testing session for all or some of the classes. In about half the ePIRLS countries, some problems occured when running the ePIRLS Software. These included computers freezing during the testing session, students preferring the keypad mouse instead of the suggested external mouse, USBs failing if used multiple times, and data upload being too slow. The release of Windows 10 just before the administration of the assessment also led to a number of issues. In all but a few cases, ePIRLS was successfully administered despite the reported issues. Most countries used individual computers and USBs to deliver ePIRLS, and five countries used both the USB method and the server method. No ePIRLS country used the server method exclusively.

National Quality Control Program

The fourth section of the Survey Activities Questionnaire addressed the National Quality Control Program that each country implemented during data collection. As part of the national quality assurance activities, National Research Coordinators were instructed to send National Quality Control Observers to ten percent of the participating schools to observe both PIRLS and ePIRLS test administration and to document compliance with the prescribed procedures. This was in addition to the program of International Quality Control visits conducted by IEA. Primarily due to budgetary constraints, some countries sent national monitors to less than ten percent of participating schools, and two countries did not send monitors to any of the testing sessions.

As shown in Exhibit 6.6, when applicable, almost all of the national centers conducted their quality assurance program using the National Quality Control Monitor Manual provided by the TIMSS & PIRLS International Study Center. Among the few documented problems detected by the national monitors were some students being late or absent, students complaining about the length of the student questionnaire, some ePIRLS technical issues, and students being confused about clicking on links during the ePIRLS testing. In addition, one case was noted where the national monitor felt the Test Administrator was unprepared.

Exhibit 6.6: Survey Activities Questionnaire, Section Four—National Quality Control Program (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Did you conduct a national quality control program that observed the data collection in the participating schools?	50	2	0
Did you use the National Quality Control Monitor (NQCM) Manual and the Classroom Observation Record provided by the TIMSS & PIRLS International Study Center to conduct your national quality control program?	48	2	0 *(Not Answered)* 2 *(Not Applicable)*
Did your national quality control monitors (NQCMs) document any major problems or special circumstances that occurred frequently during the assessment administration?	7	43	0 *(Not Answered)* 2 *(Not Applicable)*

Preparing for and Scoring the Constructed Response Items

Exhibit 6.7 provides data on responses to items asking National Research Coordinators about their experiences preparing for and scoring the constructed response items. All National Research Coordinators found the scoring procedures as explained in the *Survey Operations Procedures Unit 6—Scoring the Constructed Response Items,* including the ePIRLS supplement, to be clear and sufficient. Countries reporting problems with the scoring training materials asked for more "borderline" examples, including more detailed explanations within the scoring guides. Some countries also reported difficulties translating the examples both in the scoring guides and in the training materials. More than half of National Research Coordinators reported creating their own national examples and practice papers for training their scorers, as suggested by the TIMSS & PIRLS International Study Center.

About half of ePIRLS countries reported some minor problems using the ePIRLS Online Scoring System. The reported problems included the system being slow, the system not responding at times, issues with the "zooming function," a few student responses being assigned to scorers more than once, issues with the "flag function," and some countries wished to have a training module to be used before the actual scoring began.

Exhibit 6.7: Survey Activities Questionnaire, Section Five—Preparing for and Scoring the Constructed Response Items (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the "PIRLS 2016 Survey Operations Procedures Unit 6 – Scoring the Constructed Response Items" clear and sufficient?	52	0	0
Was the information provided in the "ePIRLS Supplement to the PIRLS 2016 Survey Operations Procedures Unit 6 – Scoring the Constructed Response Items" clear and sufficient?	15	0	0
Did you encounter any problems using the scoring training materials, provided by the TIMSS & PIRLS International Study Center?	13	39	0
Did you create national scoring training materials in addition to the international scoring training materials?	28	24	0
Did you scan the achievement booklets for electronic image scoring?	16	36	0
Did you encounter any problems using the ePIRLS Online Scoring System?	8	7	0
Did you encounter any problems during the Trend Reliability Scoring?			
Procedural problems	3	42	0 *(Not Answered)* 7 *(Not Applicable)*
Technical, software related problems	13	32	0 *(Not Answered)* 7 *(Not Applicable)*
Did all your scorers participate in scoring student responses of the trend items?	29	16	0 *(Not Answered)* 7 *(Not Applicable)*
Did you encounter any problems during the Cross-country Reliability Scoring?			
Procedural problems	2	45	0 *(Not Answered)* 5 *(Not Applicable)*
Technical, software related problems	14	33	0 *(Not Answered)* 5 *(Not Applicable)*
Did all your scorers participate in the Cross-country Reliability Scoring?	20	27	0 *(Not Answered)* 5 *(Not Applicable)*

Sixteen countries scanned their PIRLS achievement booklets and scored student responses electronically. Some technical problems were encountered while using the IEA's Coding Expert Software for the trend and cross-country scoring. Mostly countries reported the scans displayed via the Coding Expert Software being of poor quality and difficult for the scorers to read. Because

English was used for the cross-country reliability scoring task, five countries were unable to participate. For the countries that did not participate in the previous cycle of PIRLS, the question on the trend reliability scoring procedures did not apply.

Creating the Databases

The last section of the Survey Activities Questionnaire addressed data entry of the paper assessment instruments and data quality control activities. As shown in Exhibit 6.8, almost all of the National Research Coordinators found the instructions in *Survey Operations Procedures Unit 7* to be clear and sufficient. Some National Research Coordinators expressed a wish for a more automated data entry process in WinW3S, as some issues arose with the import and export functions. Also, the administration mode of the school, teacher, and home questionnaires was set to the same mode for all respondents. If some respondents, especially parents or guardians, chose to complete their questionnaire online, this status had to be adjusted manually.

Exhibit 6.8: Survey Activities Questionnaire, Section Six—Creating Databases (Numbers of NRC Responses)

Question	Yes	No	Not Answered
Was the information provided in the “PIRLS 2016 Survey Operations Procedures Unit 7 – Creating the Databases” clear and sufficient?	50	2	0
Did you encounter any problems entering test administration information and exporting your WinW3S database?	16	36	0
Who primarily entered the data for your country?			
National center staff	13	-	0
Temporarily hired data entry staff	20	-	0
An external data entry firm	8	-	0
Combination of the above	8	-	0
Other	3	-	0
Did you use manual (key) data entry to create the data files for your country?			
Achievement booklets	35	15 *(Optical Scanning)*	0 *(Not Answered)* 2 *(Not Applicable)*
Context questionnaires	39	12 *(Optical Scanning)*	0 *(Not Answered)* 1 *(Not Applicable)*
Did you encounter any problems using the IEA’s Data Manager Expert (DME) software?	4	48	0
If you entered data manually, did you enter 5% of each survey instrument twice as a quality control measure?	35	8	0 *(Not Answered)* 9 *(Not Applicable)*
Did you apply all the data quality checks described in the “PIRLS 2016 Survey Operations Procedures Unit 7 – Creating the Databases” before submitting your data to IEA Hamburg?	51	1	0
Have you stored all achievement booklets and context questionnaires in a secure storage area until the original documents can be discarded?	52	0	0

Most countries reported hiring temporary data entry staff to enter data manually. In 13 countries, the national center staff entered data from the paper instruments. A number of countries used optical scanning instead of manual data entry. All but one country reported applying all required data quality checks. All countries reported having securely stored their original assessment instruments until all data are processed and reported, and these materials can be destroyed.

CHAPTER 7

Translation and Layout Verification for PIRLS 2016

David Ebbs
Erin Wry

Introduction

The TIMSS & PIRLS International Study Center developed the international versions of the PIRLS 2016 assessment instruments, context questionnaires, and procedural manuals in English. Using the international source versions, participating countries translated the materials into their languages of instruction and adapted them to their cultural context as necessary. To ensure that the translations were of the highest quality and comparable across all of the participating countries and benchmarking entities, countries followed standard internationally agreed-upon procedures in preparing national versions of the assessment instruments (see Chapter 6: Survey Operations Procedures). The ultimate goal of the translation and adaptation process was to create national versions of the PIRLS 2016 instruments that accommodate national languages and context while maintaining international comparability.

As part of the PIRLS international quality assurance program, each country's instruments underwent a formal external review of the translations and adaptations by linguistic and assessment experts. The review included two stages: translation verification and layout verification. Translation verification was managed by IEA Amsterdam and layout verification was conducted by the TIMSS & PIRLS International Study Center. Each verification was conducted twice—once before the field test and again before the main data collection.

During translation verification, verifiers compared the national text to the international text and provided detailed feedback to improve the accuracy and comparability of the national translations. Once the verification was completed, the National Research Coordinators (NRCs) reviewed the feedback, revised their national materials as needed, and documented their changes. Following translation verification, the national instruments were sent to the TIMSS & PIRLS International Study Center for layout verification. During layout verification, verifiers checked to ensure that all national instruments conformed to the international format and that any national adaptations made to the PIRLS 2016 international instruments did not unduly influence their international comparability.

The PIRLS assessment materials that underwent verification are:

- Student achievement passages and items for PIRLS and PIRLS Literacy
- Context questionnaires for students, parents, teachers, and school principals
- Covers and directions (for each achievement booklet and context questionnaire)
- Online covers and directions (for countries administering questionnaires to parents, teacher, and/or schools online)

Countries participating in ePIRLS also submitted translated and adapted tasks and items for ePIRLS. ePIRLS utilized a comprehensive online translation system that accommodated the translation and vertification processes. All ePIRLS translations were submitted directly into the online translation system and verifiers used the system to conduct their reviews and enter their feedback.

Providing the Instruments for Translation and Adaptation

For PIRLS, the TIMSS & PIRLS International Study Center provided each country's NRC with electronic files consisting of all materials to be translated and adapted, as well as the National Adaptation Forms for documenting each step of the adaptation, translation, and verification processes. For countries that participated in ePIRLS, the TIMSS & PIRLS International Study Center provided NRCs with PDF storyboards of the ePIRLS tasks as well as documentation on accessing the ePIRLS Online Translation System.

As part of the PIRLS assessment design, most of the achievement passages appeared in several booklets, therefore the component parts of the booklets (blocks, covers, and directions) were prepared as separate files for translation and translation verification to facilitate these processes. This approach allowed countries to translate each component only once before assembling the booklets.

Following verification and approval of each country's translations, the blocks, covers, and directions were assembled into booklets to be reviewed during layout verification. To assist in this process, the TIMSS & PIRLS International Study Center provided NRCs with detailed manuals and instructional videos containing information on how to work with the electronic files, support materials for right-to-left languages, guidelines for adaptation, instructions for booklet assembly, and PDF versions of the final instruments and questionnaires for reference.

Blocks of Achievement Items Designated to Measure Trends

According to the PIRLS design, about half the passages and items are carried over from one cycle to the next for the purpose of measuring changes in student achievement over time. Accordingly, PIRLS 2016 included some passages and items previously used in PIRLS 2011 and 2006. To ensure the quality of the PIRLS trend measurement, these "trend" passages and accompanying achievement items must be administered in exactly the same way in every cycle. For countries that previously participated in PIRLS 2011 or PIRLS 2006, the PIRLS 2016 trend materials were reviewed during translation and layout verification in comparison with those from the last cycle in which the country participated. Any deviations from the previous cycle were documented by the verifiers. If a country determined that changes to an item in a trend block were absolutely necessary (e.g., in order to correct a mistranslation discovered in a previous version), they were instructed to document the change for further review during the verification process. A trend item that underwent changes was not included in the scaling process or the estimation of the achievement scores for that country.

The National Adaptations Forms

Each country prepares one National Adaptations Forms (NAFs) for each set of PIRLS assessment instruments and questionnaires for each language in which they administer the assessment and questionnaires. NAFs are Excel documents formatted to contain the complete translation, adaptation, and verification history of each set of national instruments.

When countries translated and adapted their national PIRLS and PIRLS Literacy instruments, the NAFs were filled out by the translators, reviewers, and NRCs. The translator and reviewer documented the initial adaptations made to the instruments, which the NRCs then reviewed and consolidated. Once the NAFs were updated and revised, they were reviewed again during layout verification. NRCs were responsible for responding and updating the documentation within the NAFs after each round of international verification.

Documenting an adaptation in the NAFs required entering the identifying information (location and/or question number), an English back translation of the adaptation, and recoding instructions (if applicable). To ease the process of documentation and review, NAFs include designated areas for each respondent to comment on each item within each instrument.

For ePIRLS, NAFs were not external worksheets but built into the ePIRLS Online Translation System. All national adaptations and documentation for the ePIRLS instruments were recorded directly into the Translation System, and the system has a function to export all documentation including translations, adaptations, and comments from the translators, verifiers, and NRCs.

Guidelines for Translation and Adaptation

The TIMSS & PIRLS International Study Center provided guidelines for translating and adapting the PIRLS assessment instruments. The purpose of the guidelines was to ensure that, when countries translated and adapted the international versions, the meaning and difficulty level of the instruments remained the same. All participating countries were expected to follow these guidelines, including countries that administer the instruments in English. English-speaking countries were required to adapt the international text to their national contexts to conform with English usage in the country.

In accordance with the guidelines, translators and reviewers ensured that:

- The translation is at an appropriate level for the target population
- No information is omitted, added, or clarified in the translated text
- The translated text has the same meaning as the international version and uses equivalent terminology
- The translated text has the same register (language level and degree of formality) and level of difficulty as the international version
- Idiomatic expressions are translated appropriately, not necessarily word for word
- The translated text uses correct grammar, punctuation, qualifiers, and modifiers, as appropriate for the target language

After the field test, the TIMSS & PIRLS International Study Center provided NRCs with a list of changes to the international version that they could refer to while preparing their assessment instruments for the main data collection. This information minimizes the translation burden while highlighting the necessary change to the translation before the assessment.

The Target Language

For many countries, identifying the language of assessment, referred to as the "target" language, was relatively straightforward because there is a primary language used in the education system. Other countries use more than one language of instruction in their education systems, and in these cases they translated the PIRLS instruments into multiple languages. These multilingual countries also translated the context questionnaires and administration scripts for each language assessed, and some countries also translated the home questionnaire into additional languages in order to make the questionnaire more accessible to parents from different backgrounds.

Scope of Translation and Layout Verification in PIRLS 2016

Exhibits 7.1 through Exhibit 7.3 show the languages utilized for the PIRLS, PIRLS Literacy, and ePIRLS assessments. 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.[1] Of these participants, 24 countries and 4 benchmarking entities administered the instruments in more than one language.

Exhibit 7.1: Languages used for the PIRLS 2016 Assessment Instruments

Country	Language	Instruments				
		Achievement Test	Student Questionnaire	Home Questionnaire	Teacher Questionnaire	School Questionnaire
Australia	English	●	●	●	●	●
Austria	German	●	●	●	●	●
Azerbaijan	Azerbaijani	●	●	●	●	●
	Russian	●	●	●	●	●
Bahrain	Arabic	●	●	●	●	●
	English	●	●	●	●	●
Belgium (Flemish)	Dutch	●	●	●	●	●
Belgium (French)	French	●	●	●	●	●
Bulgaria	Bulgarian	●	●	●	●	●
Canada	English	●	●	●	●	●
	French	●	●	●	●	●
Chile	Spanish	●	●	●	●	●
Chinese Taipei	Traditional Chinese	●	●	●	●	●
Czech Republic	Czech	●	●	●	●	●
Denmark	Danish	●	●	●	●	●
England	English	●	●	●	●	●
Finland	Finnish	●	●	●	●	●
	Swedish	●	●	●	●	●
France	French	●	●	●	●	●
Georgia	Georgian	●	●	●	●	●
	Azerbaijani	●	●	●	●	●
Germany	German	●	●	●	●	●

1 Counts may be inconsistent with Exhibits 7.1, 7.2, and 7.3 due to omission of benchmarking entities that share instruments with the national country participant and did not require additional translation and layout verification.

Exhibit 7.1: Languages used for the PIRLS 2016 Assessment Instruments (Continued)

Country	Language	Instruments				
		Achievement Test	Student Questionnaire	Home Questionnaire	Teacher Questionnaire	School Questionnaire
Hong Kong SAR	Traditional Chinese	●	●	●	●	●
Hungary	Hungarian	●	●	●	●	●
Iran, Islamic Rep. of	Farsi	●	●	●	●	●
Ireland	English	●	●	●	●	●
	Irish		●	●	●	●
Israel	Arabic	●	●	●	●	●
	Hebrew	●	●	●	●	●
Italy	Italian	●	●	●	●	●
Kazakhstan	Kazakh	●	●	●	●	●
	Russian	●	●	●	●	●
Latvia	Latvian	●	●	●	●	●
	Russian	●	●	●	●	●
Lithuania	Lithuanian	●	●	●	●	●
	Russian	●	●			
	Polish	●	●			
Macao SAR	Traditional Chinese	●	●	●	●	●
	English	●	●	●	●	●
	Portuguese	●	●	●	●	●
Malta	Maltese	●	●	●		
	English				●	●
Morocco	Arabic	●	●	●	●	●
Netherlands	Dutch	●	●	●	●	●
New Zealand	English	●	●	●	●	●
	Maori	●	●	●	●	●
Northern Ireland	English	●	●	●	●	●
Norway	Bokmål	●	●	●	●	●
	Nynorsk	●	●	●	●	●
Oman	Arabic	●	●	●	●	●
	English	●	●	●	●	●
Poland	Polish	●	●	●	●	●
Portugal	Portuguese	●	●	●	●	●
Qatar	Arabic	●	●	●	●	●
	English	●	●	●	●	●

Exhibit 7.1: Languages used for the PIRLS 2016 Assessment Instruments (Continued)

Country	Language	Instruments				
		Achievement Test	Student Questionnaire	Home Questionnaire	Teacher Questionnaire	School Questionnaire
Russian Federation	Russian	●	●	●	●	●
Saudi Arabia	Arabic	●	●	●	●	●
	English	●	●	●	●	●
Singapore	English	●	●	●	●	●
	Chinese			●		
	Tamil			●		
	Malay			●		
Slovak Republic	Hungarian	●	●	●	●	●
	Slovak	●	●	●	●	●
Slovenia	Slovene	●	●	●	●	●
Spain	Spanish	●	●	●	●	●
	Galician	●	●	●	●	●
	Valencian	●	●	●	●	●
	Basque	●	●	●	●	●
	Catalan	●	●	●	●	●
	English	●				
Sweden	Swedish	●	●	●	●	●
Trinidad & Tobago	English	●	●	●	●	●
United Arab Emirates	Arabic	●	●	●	●	●
	English	●	●	●	●	●
	French (Dubai only)	●	●	●	●	●
United States	English	●	●	●	●	●
Benchmarking Participants						
Buenos Aires, Argentina	Spanish	●	●	●	●	●
Eng/Afr/Zulu – RSA (5)	Afrikaans	●	●	●	●	●
	English	●	●	●	●	●
	IsiZulu	●	●	●	●	●

Exhibit 7.2: Languages used for the PIRLS Literacy 2016 Assessment Instruments

Country	Language	Instruments				
		Achievement Test	Student Questionnaire	Home Questionnaire	Teacher Questionnaire	School Questionnaire
Egypt	Arabic	●	●	●	●	●
Iran, Islamic Rep.of	Farsi	●	●	●	●	●
Kuwait	Arabic	●	●	●	●	●
	English (US)	●	●	●	●	●
	English (UK)	●	●	●	●	●
Morocco	Arabic	●	●	●	●	●
South Africa	Afrikaans	●				
	English	●				
	IsiZulu	●				
	Setswana	●	●	●	●	●
	Sesotho	●	●	●	●	●
	Sepedi	●	●	●	●	●
	isiXhosa	●	●	●	●	●
Benchmarking Participants						
Denmark (3)	Danish	●	●	●	●	●

Exhibit 7.3: Languages used for the ePIRLS 2016 Assessment Instruments

Country	Language
Canada	English
	French
Chinese Taipei	Traditional Chinese
Denmark	Danish
Georgia	Georgian
	Azerbaijani
Ireland	English
Israel	Arabic
	Hebrew
Italy	Italian
Norway (5)	Bokmal
	Nynorsk
Portugal	Portuguese
Singapore	English
Slovenia	Slovene
Sweden	Swedish
United Arab Emirates	Arabic
	English
United States	English
Benchmarking Participants	
Abu Dhabi, UAE	Arabic
	English
Dubai, UAE	French

Translators and Reviewers

All countries and benchmarking participants were advised to hire highly qualified translators and reviewers well suited to the task of working with the PIRLS materials.

Essential qualifications for translators and reviewers included:

- Excellent knowledge of English
- Excellent knowledge of the target language
- Experience of the country's cultural context
- Experience in translating literary texts, preferably at the level of the target grade

The primary responsibility of the reviewer was assessing the readability and accuracy of the translation for the target population. In addition to excellent language skills and knowledge of the country's cultural context, reviewers were expected to have experience with students in the target grade (preferably as a school teacher).

In cases where several translators and reviewers were needed for each language to distribute the work, NRCs were responsible for maintaining the consistency of the translations within and across instruments. When countries administer the assessment in more than one language, the NRCs were advised to employ translators and reviewers highly proficient in the various languages to ensure the consistency of the translations and adaptations across the different language versions.

Translation and Adaptation of the Achievement Test

When translating the PIRLS achievement passages and items, one of the main challenges is finding appropriate terms and expressions in the target language(s) that convey the same meaning and style of text as the international version. When adapting and translating expressions with more contextually appropriate terms or phrases, translators ensured that the meaning and difficulty of the passage or item remained the same as the international version. For example, it was important that adaptation/translation does not simplify or clarify the text in such a way as to provide a hint or definition of the meaning of a question. Translators also ensured the consistency of adaptations and translations from item to item. For multiple choice items, translators were instructed to pay particular attention to the literal and synonymous matches of text in both the question stem and answer options; matches in the international version were required to be maintained in the translated national version.

Although NRCs were strongly advised to keep adaptations to a minimum, some adaptations were necessary in order to prevent students from facing unfamiliar contexts or vocabulary that could hinder their ability to read and understand the passage or item. In some cases, changes to the instruments were necessary to follow national conventions of measurement, punctuation, and expressions of date and time. For example, a reference to the working week as Monday to Friday might be adapted according to national customs; similarly, a word such as "flashlight" in American English would be adapted to "torch" in British English. In addition, fictional names of characters and places were modified to similar names in the target language. When adapting the names of fictional cities or towns, translators were advised against using real names of places to prevent student responses' from being influenced by their perceptions and knowledge of the real locations.

Within the PIRLS text, some terms could not be adapted or changed beyond translation. Examples included proper names of actual people and places. To aid in the standardization of the most common adaptations across countries, the TIMSS & PIRLS International Study Center provided a list of specific examples of acceptable and unacceptable adaptations, including a list of measurement conversions.

Translation and Adaptation of the Questionnaires

Translation procedures for the questionnaires differed from the assessment passages and items in that participating countries were required to adapt some terms to ensure that questions were appropriate for the national context and education system. The terms requiring adaptation were listed in angle brackets in the international version with a description of what country-specific information was needed. For example, <language of test> and <fourth grade> would be adapted to the actual language and grade in which the assessment is administered—in the Netherlands, these terms would be replaced by equivalents "Nederlands" (Dutch) and "groep 6" (grade 4).

The guidelines for translation and adaptation contained detailed descriptions of the questionnaire adaptations, including the intent of each required adaptation, to clarify the meaning of the terms used and to enable the translators to select the appropriate national term or expression to convey the intended meaning. For PIRLS 2016, the main difficulties encountered in adapting the questionnaires involved terminology, specific educational contexts, and, for a few countries, consistency across multiple languages of administration.

Countries were permitted to add a limited number of questions to the questionnaires that were of national interest. To avoid influencing responses to the international questions, NRCs were advised to place these national questions at the end of the corresponding module or questionnaire and to ensure these questions adopt the same format as the rest of the questionnaire. The inclusion of national questions in the final questionnaires were required to be approved by the TIMSS & PIRLS International Study Center during Layout Verification.

International Translation Verification

The national translations of the international instruments were required to undergo international translation verification. IEA Amsterdam managed the international translation verification process in coordination with external translation verification companies—for PIRLS, cApStAn Linguistic

Quality Control (based in Brussels, Belgium) and for ePIRLS, EasyTranslate (based in Copenhagen, Denmark).

The required qualifications for verifiers included:

- Fluency in English
- Mother tongue proficiency in the target language
- Formal credentials as translators working in English
- University-level education and (if possible) familiarity with the subject area
- Residency in the target country, or close contact with the country and its culture

The IEA trained all international translation verifiers and supplied them with a comprehensive set of instructional materials to support their work. For PIRLS 2016, verifiers were trained through web-based seminars and were provided with information about PIRLS and the assessment instruments. Each verifier received a document containing the translation and adaptation guidelines, relevant manuals and instruments, and an instructional document containing the directions and guidelines for reviewing the national instruments and documenting deviations from the international version.

The Translation Verification Process

The instruction and training given to the verifiers emphasized the importance of maintaining the same meaning and difficulty level in the translations and adaptations as in the international versions and ensuring that translations and adaptations were adequate and consistent within and across national instruments. The translation verification process involved:

- Checking the accuracy, linguistic correctness, and comparability of the translation and adaptations of the achievement items and questionnaires
- Documenting any deviations between the national and international versions, including additions, deletions, and mistranslations
- Suggesting an alternative translation/adaptation to improve the accuracy and comparability of the national instruments

For PIRLS 2016, verifiers provided feedback from translation verification in both the sets of instruments and the associated NAFs and were asked to correct the text of the assessment items and questionnaires and to add comments describing the errors. For ePIRLS the verifiers were able to edit the text and add comments by using the ePIRLS Online Translation System.

During translation verification, some of the typical errors identified by the verifiers included typographical and grammatical errors, omissions/additions of text, mistranslations, adaptations of names (fictional versus real), gender agreement issues, and inconsistent translations (literal versus synonymous matches). After reviewing the documented comments and suggestions from the verifiers, NRCs were able revise and improve their national versions.

The translation verifiers were also instructed to document any discrepancies found in the trend items in the NAFs. Upon completion of the translation verification process, NRCs were advised to carefully review all discrepancies and to discuss any documented changes to the trend passages with the TIMSS & PIRLS International Study Center.

All comments from the verifiers included a description of the adaptation or a suggestion for revision and a code indicating the severity of the change (see Exhibit 7.4). The code was assigned by the verifier to help the NRC prioritize the necessity of each suggested revision. Comments from the verifiers that indicated major deviations, national adaptations, or incorrect adaptations were documented in the NAFs for review by the NRC and the TIMSS & PIRLS International Study Center.

Exhibit 7.4: Verification Feedback Codes for PIRLS 2016

The criteria for coding are as follows:	
CODE 1 indicates a major change or error. Examples include the omission or addition of a question or answer option; incorrect translation that changes the meaning or difficulty of the item or question; and incorrect order of questions or answer options in a multiple-choice question. If in any doubt, verifiers are instructed to use **CODE 1?** so that the error can be referred to the TIMSS & PIRLS International Study Center for further consultation	**CODE 2** indicates a minor change or error, such as a spelling or grammar error that does not affect comprehension. **CODE 3** indicates that while the translation is adequate, the verifier has a suggestion for an alternative wording. **CODE 4** indicates that an adaptation is acceptable and appropriate.

Layout Verification

Following translation verification, all national instruments were required to undergo layout verification by the TIMSS & PIRLS International Study Center. Layout verification is the final external review and ratification of each participating country's assessment instruments, questionnaires, and corresponding National Adaptations Forms. During the layout verification process, staff at the TIMSS & PIRLS International Study Center reviewed all national instruments to ensure international comparability of overall layout and proper documentation of any and all adaptations.

In particular, layout verification focused on the following:

- Reviewing the national assessment instruments for acceptable layout structure including pagination, page breaks, item sequence, response options, text formats, and graphics
- Reviewing the national adaptations applied to both the international achievement booklets and context questionnaires with respect to how they may influence the international comparability of the data

Layout Verification of Achievement Booklets and ePIRLS Tasks

The primary goal of layout verification of achievement materials is to ensure that students in different countries experience the assessment instruments in the same way. Thus, the PIRLS and PIRLS Literacy national achievement booklets were checked against the international versions to identify any deviations from the international format. Similarly, the national ePIRLS tasks were checked in comparison to the international ePIRLS tasks to detect any deviations that may interfere with the assessment.

Due to differences in languages, the PIRLS and PIRLS Literacy national assessment instruments varied slightly in length and format across countries. The international versions, however, were designed with this in mind, and extra space was provided in the margins of the pages to facilitate the use of longer text and different paper sizes (letter versus A4) without necessitating extensive changes to the layout of each page. For ePIRLS, the length of the assessment tasks remained the same for all countries but differences between languages did result in minor spacing issues for some tasks.

In addition to reviewing the overall layout of each page, verifiers also checked for proper implementation of headers, footers, section titles, graphics and number of scoring boxes displayed for each item. This included a careful review of all right to left languages to ensure that no elements of the assessment were incorrectly altered in adjusting the layout to a right to left alignment. Any layout deviations or errors, as well as any concerns of international incomparability of assessment items, were documented by the verifiers in the NAFs.

Following layout verification, the NAFs containing the verifiers' comments were sent back to the NRCs for consideration. The National Research Coordinators were asked to confirm that each of the suggested changes was implemented or provide an explanation for not implementing the suggested change.

Layout Verification of Context Questionnaires

As with the achievement booklets, the context questionnaires were checked against the international versions to identify any potential layout issues as well as to ensure the international comparability of the questionnaire data. During the layout verification, the verifiers took into consideration any national adaptations documented by the NRCs. Instances of internationally incomparable adaptations or errors were recorded by the verifiers in the NAFs along with recommendations for recoding or rewording.

In an effort to make the questionnaires general enough for international analyses but appropriate for each intended audience, participating countries were required to adapt certain phrases and designations in the text of the questionnaires. For example, items asking about levels of education were expressed in terms of the current version of the International Standard

Classification of Education (ISCED) system, ISCED 2011 (UNESCO Institute for Statistics, 2012), and required adaptation to the nationally equivalent educational terms by each participating country. During layout verification these items were reviewed in comparison to the ISCED level classifications, and if deemed internationally comparable, suggestions were made by the verifier to revise or recode their education categories.

Additionally, the verifiers ensured that all items requiring adaptations were accompanied by proper English back translations. The documentation for these universally adapted questionnaire items was intended for later use in the National Adaptations Database. The database is a compilation of each country's intended adaptations, to be used during data processing by IEA Hamburg (see Chapter 9), and the information included in the database is reported as a supplement to the *User Guide for the PIRLS 2016 International Database*.

Outcomes and Summary for PIRLS 2016

As with previous cycles of PIRLS, PIRLS 2016 incorporated stringent procedures for translation, adaptation, and verification to ensure the production of high quality and internationally comparable national instruments. In addition to the manuals and documents for instrument preparation, the TIMSS & PIRLS International Study Center provided NRCs with comprehensive guidelines about their responsibilities, from appointing highly skilled and experienced translators, to ensuring the accuracy of the documentation of national adaptations recorded in the NAFs, and responding to feedback from the verifications.

During translation and verification procedures for PIRLS 2016, translation verifiers made comments and suggestions on the following types of errors: typographical, grammar, omissions/additions of text, mistranslations, adaptations of names (fictional versus real), gender agreement, and inconsistent translations (literal versus synonymous matches, adaptation of ISCED levels). The translation verification feedback helped NRCs to improve the quality and comparability of their national instruments. Similarly, the feedback from the layout verification provided NRCs with explanations for the adjustments requested and helped ensure the international comparability of instruments across countries. From the verification feedback and outcomes to the thorough documentation of national adaptations in the NAFs, the results indicate that countries followed the rigorous guidelines, policies, and procedures in producing high quality national instruments for PIRLS 2016.

Reference

UNESCO Institute for Statistics. (2012). *ISCED: International standard classification of education*. Retrieved January 20, 2016, from http://uis.unesco.org/en/topic/international-standard-classification-education-isced.

CHAPTER 8

Quality Assurance Program for PIRLS 2016

Ieva Johansone
Erin Wry

Considerable efforts were made to develop standardized materials and survey operations procedures so that the PIRLS 2016 data met the highest standards. To document data collection activities and verify that the standardized procedures were followed, the TIMSS & PIRLS International Study Center working with IEA Amsterdam developed and implemented an International Quality Assurance Program, whereby International Quality Control Monitors visited a sample of schools in each country and observed the PIRLS 2016 administration. The purpose of this chapter is to provide an overview of the International Quality Assurance Program and report on the data collected through the program.

Overview

The International Quality Assurance Program was implemented by independent International Quality Control Monitors (IQCMs) appointed by IEA Amsterdam. The major task of the IQCMs was to conduct site visits during the data collection process. In each country, the IQCM visited a sample of 15 participating schools during the assessment administration. When there were one or more benchmarking participants from the same country and only one centrally organized national center responsible for all aspects of data collection, the IQCM visited five additional schools in each benchmarking entity on top of the schools visited for the country as a whole. For countries participating in ePIRLS, the IQCM made two visits per school—the first visit to observe the PIRLS testing session, and then another visit to observe the ePIRLS testing session and interview the School Coordinator responsible for overseeing the survey operations for that school.

In each school that they visited, IQCMs observed the PIRLS testing sessions and recorded their observations, noting any deviations from the standardized administration script, timing, and procedures, and interviewed the School Coordinators about their experiences coordinating the assessment. In addition, IQCMs checked whether the suggestions made by the international translation and layout verifiers for the national achievement booklets and context questionnaires had been integrated into the final assessment instruments, as documented in the National Adaptations Forms. Since ePIRLS translation and verification were conducted via the ePIRLS Online Translation System, all changes were tracked by the software and no additional IQCM checking was necessary for ePIRLS instruments.

Prior to beginning their assignments, the IQCMs attended a mandatory training session conducted by the TIMSS & PIRLS International Study Center. There were two training sessions, one for Southern Hemisphere countries (September) and one for Northern Hemisphere countries (January). During the training, IQCMs were introduced to the PIRLS and ePIRLS survey operations procedures and the design of the PIRLS/PIRLS Literacy 2016 achievement booklets, ePIRLS assessment tasks, and context questionnaires. IQCMs were also supplied with a manual detailing their role and responsibilities as well as the necessary materials for completing the quality control tasks.

An important aspect of the International Quality Assurance Program is the independence of the IQCMs from the national centers. In most participating countries and benchmarking entities, IEA Amsterdam recruited IQCMs who had served in the same role in previous IEA assessments. For the remaining countries, National Research Coordinators assisted IEA Amsterdam in nominating an International Quality Control Monitor. The nominated person could not be a member of the national center, or a family member or personal friend of the National Research Coordinator. Often, this person was a school inspector, ministry official, or retired school teacher. The IQCM was required to be fluent in both English and the language(s) spoken in the country.

When necessary, the IQCMs were permitted to recruit assistants to effectively cover the territory and testing timetable. For PIRLS 2016, a total of 54 IQCMs were trained across the 50 participating countries and 10 benchmarking participants. In addition, the IQCMs trained more than 200 assistant monitors. Altogether, International Quality Control Monitors observed 814 PIRLS/PIRLS Literacy testing sessions and 209 ePIRLS testing sessions. The results of the PIRLS 2016 IQCM observations are reported in the following sections of this chapter.

Quality Control Observations of the PIRLS 2016 Data Collection

International Quality Control Monitors conducted site visits during the assessment administration to a sample of 15 schools in each country. For each school visit, the IQCMs completed the Classroom Observation Record. The records were completed online via the IEA's Online SurveySystem (OSS).

The observation records were organized into the following sections:

- Section A—Documentation of the PIRLS/PIRLS Literacy Testing Session
- Section B—Summary Observations of the PIRLS/PIRLS Literacy Testing Session

- Section C—Student Questionnaire Administration and Distribution of the Learning to Read Survey
- Section D—Documentation of the ePIRLS Testing Session
- Section E—Summary Observations of the ePIRLS Testing Session
- Section F—Interview with the School Coordinator

Only IQCMs in countries participating in ePIRLS were administered Sections D and E.

Documentation and Summary Observations of the PIRLS 2016 Testing Sessions

Sections A, B, D, and E of the Classroom Observation Record addressed activities that took place during the testing sessions. The assessments were administered in two parts with a break of up to 30 minutes between each part. During test administration, IQCMs were asked to observe the activities of the Test Administrator, such as distributing, collecting, and securing the testing materials, following the assessment administration script, and timing the testing sessions.

Exhibit 8.1 reflects percentages of IQCM responses on these activities for PIRLS/PIRLS Literacy testing sessions and Exhibit 8.2 reflects this information for ePIRLS. IQCMs reported that the assessments were conducted in accordance with the international procedures. In those sessions where the total testing time for a part of the PIRLS/PIRLS Literacy assessment administration was not equal to the time allowed, it was usually because students completed their work a few minutes before the allotted time had elapsed. If Test Administrators observed students working faster than expected, a remaining-time announcement was made prior to the planned 5 minute warning to inform students that they still had ample time to complete their work. Sometimes, the break exceeded 30 minutes, and this often occurred when schools decided to follow their regular break schedule. These extended breaks were usually reported to be 35 to 45 minutes in duration.

Exhibit 8.1: Observations of PIRLS/PIRLS Literacy 2016 Assessment Administration Sessions – 814 Sessions (Percent of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did the Test Administrator distribute the PIRLS/PIRLS Literacy booklets according to the booklet assignment on the Student Tracking Form and booklet labels?	94	3	3
Did the total testing time for Part 1 of the testing session equal the time allowed?	83	15	2
Did the Test Administrator announce "you have 5 minutes left" prior to the end of Part 1 of the testing session?	83	14	3
Were there any other time remaining announcements made during Part 1 of the testing session?	12	88	0
Was the total time for the break between Part 1 and Part 2 of the testing session equal to or less than 30 minutes?	78	10	12
Were the booklets left unattended or unsecured during the break?	2	93	5
Did the total testing time for Part 2 of the testing session equal the time allowed?	83	16	1
Did the Test Administrator announce "you have 5 minutes left" prior to the end of Part 2 of the testing session?	86	13	1
Were there any other time remaining announcements made during Part 2 of the testing session?	10	90	0
Did any students finish either Part 1 or Part 2 of the PIRLS/PIRLS Literacy assessment early (before the 40 minutes were up)?	77	21	2
Did the test administrator have a watch with a seconds hand (or stopwatch/timer) for accurately timing the testing session?	90	9	1
Were the booklets collected and secured after the testing session?	91	8	1

For ePIRLS, Test Administrators ensured that the achievement test booklets were distributed and students were logged into the ePIRLS Software with their student IDs and passwords according to the Student Tracking Forms[1] and labels. In accordance with the procedure, at the end of the testing session, Test Administrators were asked to collect and secure the test booklets. The IQCMs reported that in 91 percent of the PIRLS/PIRLS Literacy testing sessions this occurred. However, in a few cases, the Student Questionnaire was attached to the test booklet, and in these cases students retained their test booklets until they completed their questionnaire.

1 As described in more detail Chapter 6, the Test Administrators used the student tracking form to verify the assignment of survey instruments to students and to indicate participation status.

Exhibit 8.2: Observations of ePIRLS 2016 Assessment Administration Sessions – 209 Sessions (Percent of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did the Test Administrator make sure that students were seated at their assigned computers (logged into the ePIRLS Software with his/her Student ID and password) according to the Student Tracking Form?	79	0	21
Did the Test Administrator read the directions (presented on each student's computer) aloud to the students?	35	50 *(students followed through the directions by themselves)*	15
Did the Test Administrator announce "you have 5 minutes left" prior to the end of Part 1 of the testing session?	55	27	18
Were there any other time remaining announcements made during Part 1 of the testing session?	6	82	12
Was the total time for the break between Part 1 and Part 2 of the testing session equal to or less than 30 minutes?	46	10	44
Were the computers and USB sticks kept secure during the break?	48	4	48
Did the Test Administrator announce "you have 5 minutes left" prior to the end of Part 2 of the testing session?	48	33	19
Were there any other time remaining announcements made during Part 2 of the testing session?	4	83	13
Did any students finish either Part 1 or Part 2 of the ePIRLS assessment early (before the 40 minutes were up)?	91	5	4

Exhibits 8.3 and 8.4 report on the activities conducted during the assessment sessions for PIRLS/PIRLS Literacy and ePIRLS, respectively. To standardize test administration, all Test Administrators were instructed to read the script in the Test Administrator Manual to the students. IQCMs reported that in more than half of the PIRLS/PIRLS Literacy observations, the Test Administrators followed the script exactly. When the Test Administrator deviated from the script, nearly all modifications were reported to be "minor." For ePIRLS, students were allowed to click through the directions on their own, and for this reason many ICQMS did not answer this question or marked it "not applicable."

Exhibit 8.3: Test Administrators Following the Test Administration Script – 814 PIRLS/PIRLS Literacy Sessions (Percent of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Had the test administrator familiarized himself or herself with the test administration script prior to the testing?	88	6	5 *(I Cannot Answer)* 1 *(Not Answered)*
Did the test administrator follow the test administration script in the PIRLS 2016 Test Administrator Manual?	58	36 *(Minor changes)* 5 *(Major changes)*	1
If the Test Administrator made changes to the script, how would you describe them?			
Additions	3	6	0 *(Not Answered)* 91 *(Not Applicable)*
Revisions	2	7	0 *(Not Answered)* 91 *(Not Applicable)*
Deletions	1	8	0 *(Not Answered)* 91 *(Not Applicable)*
Did the test administrator address student questions appropriately?	91	4	5

Exhibit 8.4: Test Administrators Following the Test Administration Script – 209 ePIRLS Sessions (Percent of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Had the test administrator familiarized himself or herself with the test administration script prior to the testing?	68	4	9 *(I Cannot Answer)* 18 *(Not Answered)*
Did the test administrator follow the test administration script in the PIRLS 2016 Test Administrator Manual?	41	38 *(Minor changes)* 5 *(Major changes)*	16
If the Test Administrator made changes to the script, how would you describe them?			
Additions	4	7	3 *(Not Answered)* 86 *(Not Applicable)*
Revisions	3	7	4 *(Not Answered)* 86 *(Not Applicable)*
Deletions	2	10	2 *(Not Answered)* 86 *(Not Applicable)*
Did the test administrator address student questions appropriately?	95	2	3

Exhibit 8.5 summarizes observations on student compliance with instructions and overall cooperation during assessment administration for both PIRLS/PIRLS Literacy and ePIRLS. According to the IQCMs' observations, in almost all of the sessions, students complied well or very well with the instruction to stop work at the end of both Part 1 and Part 2 of the PIRLS/PIRLS Literacy testing sessions. In addition, IQCMs described the students as extremely orderly and cooperative during most of the testing sessions.

Exhibit 8.5: Student Cooperation During Assessment Administration – 814 PIRLS/PIRLS Literacy Sessions and 209 ePIRLS Sessions (Percent of IQCM Responses)

Question	Very Well (%)	Fairly Well (%)	Not well at all (%)	Not Answered or Not Applicable (%)
When the Test Administrator ended Part 1 of the PIRLS/PIRLS Literacy testing session, how well did the student comply with the instruction to stop work?	85	11	2	2
When the Test Administrator ended Part 2, of the PIRLS/PIRLS Literacy testing session, how well did the student comply with the instruction to stop work?	85	14	0	1

Question	Extremely (%)	Moderately (%)	Somewhat (%)	Hardly (%)	Not Answered or Not Applicable (%)
To what extent would you describe the students as orderly and cooperative during the PIRLS/PIRLS Literacy testing session?	61	31	6	1	1
To what extent would you describe the students as orderly and cooperative during the ePIRLS testing session?	74	20	2	2	2

Summary Observations of the PIRLS 2016 Testing Sessions

Exhibit 8.6 reports on the IQCMs' general observations of the PIRLS/PIRLS Literacy assessment administrations and Exhibit 8.7 reports on the IQCM's general observations of the ePIRLS administrations. Overall, IQCMs reported that the quality of testing sessions was good, very good, or excellent (90% for PIRLS/PIRLS Literacy and 93% for ePIRLS). In most of the testing sessions the IQCMs attended, no problems were observed, and in only 1 percent of cases for both PIRLS/PIRLS Literacy and ePIRLS did a student refuse to take the test. In addition, nearly all of the observed testing sessions took place under favorable room conditions that were suitable for students to work without distraction. The large majority of students, 96 percent for PIRLS/PIRLS Literacy and 92 percent for ePIRLS, followed the direction to store away everything, including electronic devices, for the duration of test administration. The IQCMs also reported that in 94 percent of observed testing sessions, students were seated in an arrangement that provided adequate space for students to work and not be distracted by one another.

Exhibit 8.6: **General Observations of the PIRLS/PIRLS Literacy 2016 Testing Sessions – 814 Sessions (Percent of IQCM Responses)**

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Did the student identification information on the PIRLS/PIRLS Literacy booklets correspond with the Student Tracking Form?	95	3	2
Were any defective test booklets detected and replaced?	2 *(BEFORE the testing began)* 1 *(AFTER the testing began)*	97 *(BEFORE the testing began)* 89 *(AFTER the testing began)*	1 *(BEFORE the testing began)* 11 *(AFTER the testing began)*
If any defective test booklets were replaced, did the Test Administrator replace them appropriately?	1	1	0 *(Not Answered)* 98 *(Not Applicable)*
Did any students refuse to take the test?	1	98	1
If a student refused, did the Test Administrator accurately follow the instructions for excusing the student?	0	0	1 *(Not Answered)* 99 *(Not Applicable)*
Were any late students admitted to the testing room?	10 *(BEFORE the testing began)* 9 *(AFTER the testing began)*	77 *(There were no late students)* 2 *(Late students were not admitted)*	2
Did any students leave the room for an "emergency" during the testing?	12	87	1
If a student left the room for an emergency during the testing, did the Test Administrator address the situation appropriately (collect the test booklet, and if re-admitted, return the test booklet)?	2	2	9 *(Not Answered)* 87 *(Not Applicable)*
Were there any students requiring special accommodations (e.g., students with visual or hearing impairment, Dyslexia)?	6	93	1
Did students store away everything, including all electronic devices, having only a pen or a pencil and the test booklet for the duration of the test administration?	92	6	2
During the testing session did the test administrator walk around the room to be sure students were working on the correct section of the test and/or behaving properly?	88	10	2
Were the conditions in the testing room suitable (lighting, temperature, noise, etc.) for the students to work without distractions?	89	9	2
Did the seating arrangement provide adequate space for students to work and not be distracted by each other?	94	5	1
Did you see any evidence of students attempting to cheat on the tests (e.g., by copying from a neighbor)?	3	96	1

Question	Excellent (%)	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Not Answered (%)
In general, how would you describe the overall quality of the testing session?	40	37	13	6	2	2

Exhibit 8.7: General Observations of the ePIRLS 2016 Testing Sessions – 209 Sessions (Percent of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Were any defective USB sticks detected and replaced?	7 *(BEFORE the testing began)* 2 *(AFTER the testing began)*	81 *(BEFORE the testing began)* 83 *(AFTER the testing began)*	12 *(BEFORE the testing began)* 15 *(AFTER the testing began)*
Did any students refuse to take the test?	1	96	3
If a student refused, did the Test Administrator accurately follow the instructions for excusing the student?	1	0	3 *(Not Answered)* 96 *(Not Applicable)*
Were any late students admitted to the testing room?	3 *(BEFORE the testing began)* 1 *(AFTER the testing began)*	73 *(There were no late students)* 2 *(Late students were not admitted)*	21
Did any students leave the room for an "emergency" during the testing?	13	83	4
Were there any students requiring special accommodations (e.g., students with visual or hearing impairment, Dyslexia)?	15	82	3
Did students store away everything (school books/papers and all electronic devices), having only the computer used for the ePIRLS testing session?	96	1	3
Were the conditions in the testing room suitable (lighting, temperature, noise, etc.) for the students to work without distractions?	93	3	4
Did the seating arrangement provide adequate space for students to work and not be distracted by each other?	94	3	3
Were all students in the participating class tested together in one session or on groups (multiple testing sessions due to the number of computers available)?	46	37 *(Multiple sessions)*	17
If laptops were used, did students have an external mouse available?	41	17	13 *(Not Answered)* 29 *(Not Applicable)*
If no external mouse was available, did using the laptop touchpads cause any problems?	1	30	23 *(Not Answered)* 46 *(Not Applicable)*
In addition to the Test Administrator, were there any additional personnel (e.g., School Coordinator, class teacher, an IT specialist) available during the testing session?	86	11	3
Did any technical problems occur during the testing session?	52	45	3
Did the Test Administrator submit the data from each computer students used for the ePIRLS testing session directly after the testing session?	46	38	16

Question	Excellent (%)	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Not Answered or Not Applicable (%)
In general, how would you describe the overall quality of the testing session?	52	31	10	4	0	3

Student Questionnaire Administration

Exhibit 8.8 summarizes the IQCMs' observations of the Student Questionnaire administration. IQCMs reported that in the majority of the testing sessions, the Student Questionnaires were distributed according to the Student Tracking Forms and questionnaire labels. In some cases, Test Administrators did not follow the Student Questionnaire administration script exactly. In the cases that the Test Administrator deviated from the script, the modifications were "minor" for the most part. In 15 percent of the observed testing sessions, Test Administrators read Student Questionnaire questions aloud, and in 58 percent of the sessions students answered these questions independently. It should be noted that some schools chose to administer the questionnaire on a different date than the assessment, and in these cases, IQCMs were not required to observe student questionnaire administration.

Exhibit 8.8: Student Questionnaire Administration – 814 Sessions (Percent of IQCM Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
When the test administrator read the script to end the PIRLS/PIRLS Literacy testing session followed by the Student Questionnaire administration, did the test administrator announce a break?	56	10	34 *(Not Applicable)*
Did the Test Administrator distribute the Student Questionnaires according to the Student Tracking Form and questionnaire labels?	62	4	34 *(Not Applicable)*
Did the test administrator follow the questionnaire administration script in the PIRLS 2016 Test Administrator Manual?	43	23 *(Minor changes)* 4 *(Major changes)*	0 *(Not Answered)* 30 *(Not Applicable)*
If the Test Administrator made changes to the script, how would you describe them?			
Additions	2	6	3 *(Not Answered)* 89 *(Not Applicable)*
Revisions	2	6	3 *(Not Answered)* 89 *(Not Applicable)*
Deletions	1	7	3 *(Not Answered)* 89 *(Not Applicable)*
Did the test administrator read the questions aloud to the students?	15	55 *(students answer the questions independently)*	0 *(Not Answered)* 30 *(Not Applicable)*
After the Student Questionnaire administration, did the Test Administrator distribute the Learning to Read Surveys?	12	58	0 *(Not Answered)* 30 *(Not Applicable)*
If the Learning to Read Surveys were distributed at this time, did the Test Administrator distribute them according to the PIRLS Student Tracking Form and survey labels?	4	3	4 *(Not Answered)* 89 *(Not Applicable)*

Interview with the School Coordinator

Section F was the final component of the Classroom Observation Record and involved the IQCM conducting an interview with the School Coordinator. The interview addressed issues such as the following:

- Shipment of assessment materials
- Arrangements for test administration
- Responsiveness of the national center to queries
- Necessity for make-up sessions
- Organization of classes in the school (to validate the within-school sampling procedure)

As shown in Exhibit 8.9, 91 percent of School Coordinators reported that the PIRLS administration in their school went "very well" or "satisfactorily" overall. In addition, the School Coordinators noted that that the School Coordinator Manual worked well for them and most other school staff members had positive attitudes toward PIRLS testing.

Exhibit 8.9: Interview with the School Coordinator, Overview – 814 Records (Percent of School Coordinator Responses)

Question	Very well, no problems (%)	Satisfactory, few problems (%)	Unsatisfactory, many problems (%)	Not Answered or Not Applicable (%)
Overall, how would you say the testing went in your school?	66	25	1	8

Question	Positive (%)	Neutral (%)	Negative (%)	Not Answered or Not Applicable (%)
Overall, how would you rate the attitude of the other school staff members towards the PIRLS testing?	59	33	5	3

Question	Worked well (%)	Needs improvement (%)	Not Answered or Not Applicable (%)
Overall, do you feel the School Coordinator Manual worked well for you or does it need improvement?	74	5	21

Exhibit 8.10 shows that there were only a small number of cases where components were missing from the shipments of test materials. In some cases where the School Coordinator reported not receiving all of the PIRLS materials, test materials were brought to the school on the testing day by external Test Administrators. The School Coordinators also reported that in 74 percent of the schools observed for PIRLS 2016, the national centers were responsive to the school's questions and concerns.

Exhibit 8.10: Interview with the School Coordinator, Details – 814 Records (Percent of School Coordinator Responses)

Question	Yes (%)	No (%)	Not Answered or Not Applicable (%)
Prior to the (first) testing day, did you have time to check your shipment of materials from the national center?	67	22	11
Did you receive the correct shipment of the materials as listed in your School Coordinator Manual and according to the tracking forms?	69	8	23
If no, did the national center provide the missing materials in time for the testing?	1	0	0 *(Not Answered)* 99 *(Not Applicable)*
Was the national center responsive to your questions or concerns?	74	4	22
Was the Teacher Questionnaire administered online?	17	81	2
If the Teacher Questionnaire was administered online, did the teacher(s) encounter any problems?	0	7	10 *(Not Answered)* 83 *(Not Applicable)*
Was the School Questionnaire administered online?	19	77	4
If the School Questionnaire was administered online, did the person completing it encounter any problems?	0	9	10 *(Not Answered)* 81 *(Not Applicable)*
Was the Learning to Read Survey administered online?	2	75	23
If the Learning to Read Survey was administered online, did the parents/guardians encounter any problems?	0	2	0 *(Not Answered)* 98 *(Not Applicable)*
Do you anticipate that a makeup session will be required at your school?	9	90	1
If yes, do you intend to conduct one?	2	1	6 *(Not Answered)* 91 *(Not Applicable)*
Did the students receive any special instructions, motivational talk, or incentives to prepare them for the assessment(s)?	46	47	7
Did you provide the list of classes in the tested grade to the national center?	83	12 *(Centralized database used)*	5
If there was another international assessment, would you be willing to serve as a School Coordinator?	82	13	5

In 46 percent of the visited schools, School Coordinators indicated that students were given special instructions, motivational talks, or incentives by a school official or the classroom teacher prior to testing. Only nine percent of School Coordinators anticipated needing a makeup session.

Because the sampling of classes requires a complete list of all classes in the school at the target grade, IQCMs were also asked to verify that all classes were included in the sampling process. School Coordinators were asked how many classes of the tested grade are in the school, how many were selected to participate, and whether he/she provided the list of classes to the national center. Over 80 percent of School Coordinators confirmed that they sent a complete list of classes to the national center. In 12 percent of the observed schools, centralized databases were used instead of class lists.

As a reflection of the successful planning and implementation of PIRLS 2016, 82 percent of respondents for both grades said that they would be willing to serve as a School Coordinator in future international assessments.

CHAPTER 9

Creating the PIRLS 2016 International Database

Sebastian Meyer
Mark Cockle
Milena Taneva

This chapter describes the procedures implemented by IEA Hamburg for verifying the PIRLS 2016 data and creating the PIRLS 2016 International Database (IDB).

Preparing the PIRLS 2016 International Database (IDB) and ensuring its integrity was a complex endeavor requiring extensive collaboration between IEA Hamburg, the TIMSS & PIRLS International Study Center, Statistics Canada, and the national centers of participating countries. Once the countries had created their data files and submitted them to IEA Hamburg, an exhaustive process of verification and editing known as "data cleaning" began.

Data cleaning is the process of checking data for inconsistencies and formatting the data to create a standardized output. The overriding concerns of the data cleaning process were to ensure:

- All information in the database conformed to the internationally defined data structure
- The content of all codebooks and documentation appropriately reflected national adaptations to questionnaires
- All variables used for international comparisons were in fact comparable across countries (after harmonization, where necessary)
- All institutions involved in this process applied quality control measures throughout in order to assure the quality and accuracy of the PIRLS 2016 data

IEA Hamburg was responsible for checking the data files from each country, applying standardized data cleaning rules to verify the accuracy and consistency of the data and documenting any deviations from the international file structure. Data files were created at each country's national center and reviewed prior to submission to IEA Hamburg. The National Research Coordinators (NRCs) from each country collaborated with IEA Hamburg to resolve any queries which emerged during the data cleaning process, and the NRCs checked interim versions of the national/benchmarking participant database(s) produced by IEA Hamburg.

The TIMSS & PIRLS International Study Center provided the NRCs with univariate data almanacs containing summary statistics on each variable so that the national centers could evaluate their data from an international perspective. The TIMSS & PIRLS International Study Center also scaled the achievement and background data, as documented in Chapter 12: Scaling the PIRLS 2016 Achievement Data and in Chapter 14: Creating and Interpreting the PIRLS 2016 Context Questionnaire Scales, and produced achievement scores (plausible values) and scores on the background scales. Using the Windows® Within-school Sampling Software (WinW3S)[1] database and processed response data provided by IEA Hamburg, Statistics Canada in collaboration with IEA Hamburg calculated the sampling weights, population coverage, and school and student participation rates—as documented in Chapter 3 and Chapter 5.

Data Sources

Data Entry and Verification of Paper Booklets and Questionnaires

Each national center was responsible for inputting the information collected in paper-based test booklets and questionnaires into computer data files using the IEA Data Management Expert (DME) software. The DME is a software system developed by IEA Hamburg that facilitates data entry and includes validation checks to identify inconsistencies. As a general principle, national centers were instructed to enter data for any booklet or questionnaire that contained at least one valid response, discarding unused or empty instruments.

National centers entered responses from the paper instruments into data files using a predefined international codebook. The codebook defines the structure of the data to be entered and contains information about the variable names, lengths, labels, and missing codes, as well as variable ranges for continuous measures or counts and valid values for nominal or ordinal questions.

As documented in Chapter 7: Translation and Layout Verification for PIRLS 2016, countries participating in PIRLS are expected to make national adaptations to certain questions in the international source versions of the context questionnaires (e.g., the questions about parents' education must be adapted to the national context). Countries making such adaptations were required to adapt the codebook structure to reflect the adaptations made to the national questionnaire versions before beginning the data entry process.

To ensure consistency across participating countries, the basic rule for data entry in the DME required national staff to enter data "as is" without any interpretation, correction, truncation, imputation, or cleaning.

The guiding principles for data entry included the following:

1 WinW3S is software developed by IEA Hamburg that stores participation information at school, teacher, class, and student levels in a relational database while maintaining a hierarchical ID system. The software allows users to perform all necessary within-school sampling according to the PIRLS standards, and also provides some data validation in and across these levels.

- Responses to closed response items entered as "1" if the first option was used, "2" if the second option is marked, and so on
- Responses to open response questions, for example number of students in the PIRLS class, entered "as is" even if the value is outside the originally expected range
- Responses to filter questions and filter-dependent questions entered exactly as filled in by the respondent, even if the information provided is logically inconsistent
- Non-response, ambiguous responses, responses given outside of the expected format, or conflicting responses (e.g., selection of two options in a multiple-choice question), coded as "omitted or invalid"

As each respondent ID number was entered, it was checked by the DME software for alignment with a five-digit checksum generated by WinW3S. A mistype in either the ID or the checksum resulted in an error message prompting the data entry person to check the entered values. The data verification module of DME also checked for a range of other issues such as inconsistencies in identification codes and out-of-range or otherwise invalid codes. When such issues were flagged by the software, the individuals entering the data were prompted to resolve the inconsistency or confirm that an issue existed before resuming data entry.

Double Data Entry

To check data entry reliability in participating countries, national centers were required to enter a 5 percent sample of each survey instrument (achievement booklet or questionnaire) twice by two different data entry persons (punchers). IEA Hamburg recommended that countries begin the double data entry process as early as possible during the data capture period in order to identify possible systematic misunderstandings or mishandlings of data entry rules and to initiate appropriate remedial actions—for example, retraining national center staff. Those entering the data were required to resolve discrepancies between the first and second data entries by consulting the original questionnaire and applying the international rules in a uniform way.

Although it was desirable that each and every discrepancy be resolved before submission of the complete dataset, the acceptable level of disagreement between the originally entered and double entered data was established at 1 percent or less for questionnaire data and at 0.1 percent or less level for achievement data. Values above this level required a complete re-entry of data.

The level of disagreement between the originally entered data and double entered data was evaluated by IEA Hamburg, and it was found that in general the margin of error observed for processed data was well below the required threshold.

Data from ePIRLS Administration

The ePIRLS assessment was designed to run on PCs using one of two methods: USB or server delivery. In the simpler USB delivery, a USB stick pre-loaded with the ePIRLS software was inserted into a USB port on a student's computer. The Test Administrator located and ran the program and then entered the ID and checksum (i.e., password) from the Student Tracking Form to begin the assessment. For the server delivery, a PC serving as a local server and having the ePIRLS software installed was connected to the school's Local Area Network (LAN), and the individual student PCs accessed the assessment over the LAN using a Firefox browser. Similar to the USB method, IDs and checksums from the Student Tracking Form were used to identify the student and begin the assessment, and checksums also contained information on which task is assigned to students.

For both delivery methods, the student response data were stored in a SQL-Compact database, the contents of which could be uploaded to the IEA Hamburg server immediately following the assessment, or later off-site. Following data upload, student responses to constructed response items were sent to the Online Scoring System, which almost immediately made student responses available to be allocated to scorers. Scoring took place directly on the IEA Hamburg server—allowing IEA Hamburg to monitor, in real time, the progress of scoring within countries.

Also available online to national centers was an upload monitor listing all the student records that had been uploaded to the IEA Hamburg server. In the rare cases that duplicate IDs were detected, the IDs were flagged and national centers indicated which record to keep. The data monitor also allows a list of IDs to be downloaded so that they can be used to update data availability status in WinW3S.

Data from Online Questionnaire Administration

As documented in Chapter 6: Survey Operations Procedures in PIRLS 2016, national centers had the option of administering the school, teacher, and home questionnaires online instead of, or in addition to, using paper-based questionnaires. Students participating in ePIRLS also completed a brief questionnaire following the assessment through the ePIRLS software. To ensure confidentiality, national centers provided every respondent with a letter containing individual login information along with information on how to access the online questionnaire. This login information corresponded to the ID and checksum provided from WinW3S, meaning that the identity validation step occurring at the national centers for paper-based questionnaires occurred when the respondents' logged in to the survey.

Online administration of questionnaires had a number of advantages. Because responses were collected in digital format and stored directly on the IEA Hamburg server, there was no need for data entry, reducing the workload for national centers. Also, the online system does not allow for inconsistent response patterns, meaning that the data collected had fewer inconsistencies when compared with data collected through the paper-based questionnaires. For example, if the directions ask the respondent to "Check one circle for each line," the system does not allow the respondent to check more than one response category on each line.

The PIRLS 2016 online questionnaires also include skip logic, which minimized response burden and improved data consistency. The PIRLS questionnaires have a number of questions that filter out respondents—meaning the subsequent questions are not applicable given the response to the filter question. For example, Question 9 of the school questionnaire reads: "Does your school have a school library? If yes, go to #9a, and if No, go to #10." If a respondent chooses "No," the online survey skips directly to Question 10, omitting Questions 9a and 9b. Not only does the skip logic save the respondents' time, it also results in fewer inconsistencies in the data received by IEA Hamburg.

Data Verification at the National Centers

Before sending the data to IEA Hamburg for further processing, national centers carried out mandatory validation and verification steps on all entered data and undertook corrections as necessary.

While the questionnaire data were being entered, the data manager or other staff at each national center used the information from the Tracking Forms to verify the completeness of the materials. Student participation information (e.g., whether a student participated in the assessment or was absent) was entered or imported into WinW3S.

The validation process was supported by an option in WinW3S to generate an inconsistency report. This report listed all of the types of discrepancies between variables recorded during the within-school sampling and test administration processes and made it possible to cross-check these data against data entered in the DME, the database for online respondents, and the uploaded student data on the central international server. When inconsistencies were identified, data managers were instructed to resolve the issue before final data submission to IEA Hamburg. If inconsistencies remained or the national center could not solve them, IEA Hamburg asked the center to provide documentation on these problems.

Upon submitting the validated data to IEA Hamburg, NRCs also provided extensive documentation including hard copies or electronic scans of all original Student and Teacher Tracking Forms, Student Listing Forms, and when applicable, a report on procedural activities collected as part of the online Survey Activities Questionnaire (see Chapter 6).

Cleaning the International and National Databases

Overview

To ensure the integrity of the international database, a uniform data cleaning process was followed, involving regular consultation between IEA Hamburg and the NRCs. After each country had submitted its data, codebooks, and documentation, IEA Hamburg, in collaboration with the NRCs, conducted a four-step cleaning procedure on the submitted data and documentation:

1. A structural check
2. A check of the identification (ID) variables
3. Linkage cleaning
4. Background cleaning

Data cleaning was an iterative process, with numerous iterations of the four-step cleaning procedure being implemented for each national data set. This repetition ensured that all data were properly cleaned and that any new errors that could have been introduced during the data cleaning were rectified. The cleaning process was repeated as many times as necessary until all data were made consistent and comparable. Any inconsistencies detected during the cleaning process were resolved in collaboration with national centers, and all corrections made during the cleaning process were documented in a cleaning report, produced for each country.

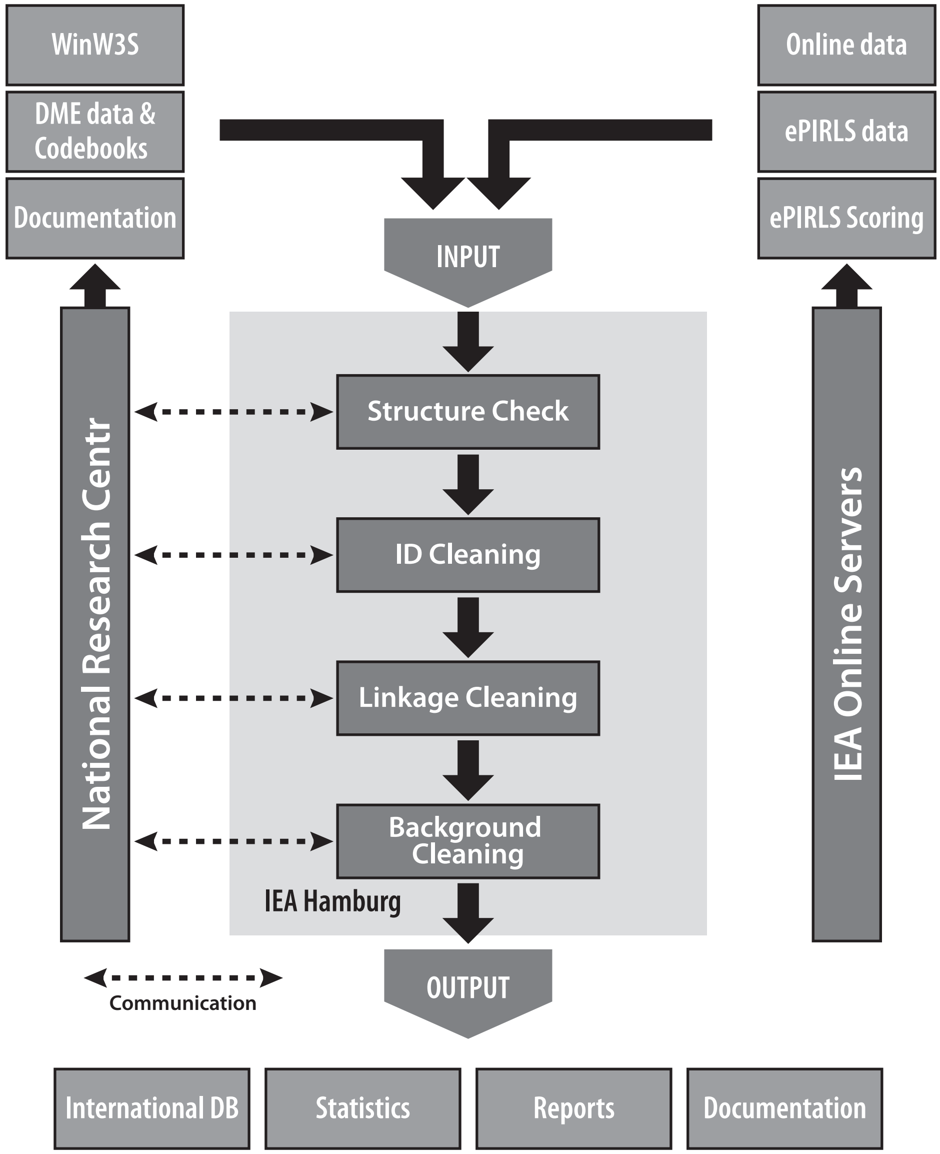
After the final cleaning iteration, each country's data were sent to Statistics Canada for the calculation of sampling weights, and then the data, including sampling weights, were sent to the TIMSS & PIRLS International Study Center so that scaling could be performed. The NRCs were provided with interim data products to review at two different points in the process.

Preparing National Data Files for Analysis

The main objectives of the data cleaning process were to ensure that the data adhered to international formats, that school, teacher, and student information could be linked across different survey files, and that the data reflected the information collected within each country in an accurate and consistent manner.

As illustrated in Exhibit 9.1, the program-based data cleaning consisted of a set of activities explained in the following subsections. IEA Hamburg carried out all of these activities in close communication with the national centers.

Exhibit 9.1: Overview of Data Processing at IEA Hamburg



Structural Check

For each country, data cleaning began with a review of data file structures and data documentation, including a review of national adaptation comments, Student Listing Forms, Student Tracking Forms, Teacher Tracking Forms, and the Survey Activities Questionnaire.

After the review, IEA Hamburg merged the tracking information and sampling information captured in the WinW3S database with the student-level databases containing the corresponding student instrument data, and, if applicable, ePIRLS data. For countries administering questionnaires through online and paper modes, IEA Hamburg merged the questionnaire data across modes of administration. At this stage, data from the different sources were transformed and imported into one SQL database so that this information would be available during all further data processing stages.

The first checks identified differences between the international and the national file structures. Some countries made adaptations (such as adding national variables or omitting or modifying international variables) to their questionnaires. The extent and nature of these changes differed across countries: some countries administered the questionnaires without any modifications (apart from translations and necessary adaptations relating to cultural or language-specific terms), whereas other countries inserted response categories within existing international variables or added national variables.

To keep track of adaptations, staff at the TIMSS & PIRLS International Study Center asked the national centers to complete National Adaptations Forms. In their adaptations, countries sometimes modified the structure and values of the international codebooks, and in these cases IEA Hamburg recoded the national data files to ensure that the resulting data remained comparable across countries. The national adaptation process is described in Chapter 7 and details about country-specific adaptations to the international instruments can be found in Supplement 2 of the *PIRLS 2016 User Guide for the International Database*.

IEA Hamburg then discarded variables created purely for verification purposes during data entry and made provisions for adding new variables necessary for analysis and reporting, including reporting variables, derived variables, sampling weights, and scale scores.

Once IEA staff had ensured that each data file matched the international format, they applied a series of standard data cleaning rules for further processing. Processing during this step employed software developed by IEA Hamburg that could identify and correct inconsistencies in the data. Each potential problem flagged at this stage was identified by a unique problem number, and then described and recorded in a database. The action taken by the cleaning program or IEA staff with respect to each problem was also recorded.

IEA Hamburg referred problems that could not be rectified automatically through the program to the responsible NRC so that national center staff could check the original data collection instruments and Tracking Forms to trace the source of the error. Wherever possible, staff at IEA Hamburg suggested a remedy and asked the national centers to either accept it or propose an alternative. If a national center could not solve the issue through verification of the instruments or forms, IEA Hamburg applied a general cleaning rule to the files to rectify the error. When all automatic updates had been applied, IEA staff used SQL recoding scripts to directly apply any remaining corrections to the data files.

Checking Identification Variables

Each record in a data file needs to have a unique identification number. The existence of records with duplicate ID numbers in a file implies an error of some kind. Some countries administered the school, teacher, and home questionnaire online in addition to on paper. This yields the possibility that a respondent could complete both the paper and the online versions of the questionnaire. Similarly, it was possible for an ePIRLS login to be used (and uploaded) twice. If two records in a PIRLS 2016 database shared the same ID number and contained exactly the same data, IEA Hamburg deleted one of the records and kept the other one in the database. In the rare case that both records contained different data and IEA staff found it impossible to identify which record contained the "true data," national centers were asked which record to keep.

Although the ID cleaning covered all data from all instruments, it focused mainly on the student files. In addition to checking the unique student ID numbers, it was crucial to check variables pertaining to student participation and exclusion status, as well as students' birth dates and dates of testing in order to calculate student age at the time of testing. The Student Tracking Forms provided an important tool for resolving anomalies in the database.

As mentioned previously, IEA Hamburg conducted all cleaning procedures in close cooperation with the national centers. After national center staff had cleaned the identification variables, they passed the clean databases with information about student participation and exclusions on to Statistics Canada, which used this information to calculate student participation rates, exclusion rates, and student sampling weights.

Cleaning Linkages

As data on students, parents, teachers, and schools appeared in a number of different data files, a process of linkage cleaning was implemented to ensure that the data files would correctly link together. The linking of the data files followed a hierarchical system of identification codes that included school, class, and student components. These codes linked the students with their class and/or school membership. Further information on linkage codes can be found in Chapter 6: Survey Operations Procedures in PIRLS 2016.

Linkage cleaning consisted of a number of checks to verify that student entries matched across achievement files, student background files, scoring reliability files, and home background files. In addition, at this stage, checks were conducted to ensure that teacher and student records linked correctly to the appropriate schools. The Student Tracking Forms, Teacher Tracking Forms, and Student Listing Forms were crucial in resolving any anomalies. IEA Hamburg also liaised with NRCs about any problematic cases, and the national centers were provided with standardized reports listing all inconsistencies identified within the data.

Background Cleaning

The amount of inconsistent and implausible responses in questionnaire data files varied considerably across countries. IEA Hamburg determined the treatment of inconsistent responses on a question-by-question basis, using all available documentation to make an informed decision. IEA Hamburg staff also checked all questionnaire data for consistency across the responses given. For example, Question 1 in the school questionnaire asked for the total school enrollment in all grades, and Question 2 asked for the enrollment in the target grade only. Logically, the number given as a response to Question 2 could not exceed the number provided by school principals in Question 1. Similarly, it is not possible that the amount of years a teacher has been teaching altogether (Question 1 in the teacher questionnaire) exceeds his/her age (Question 3 in the teacher questionnaire). IEA Hamburg flagged inconsistencies of this kind and then asked the national centers to review these issues. IEA staff recoded those cases that could not be corrected as "invalid."

Filter questions, which appeared in some questionnaires, directed respondents to a particular subquestion. IEA Hamburg applied the following cleaning rule to these filter questions and the dependent questions that followed: If a respondent answered "No" to Question 9 in the school questionnaire "Does your school have a school library?," IEA Hamburg recoded any responses to the dependent questions as "logically not applicable." Also, following the same example, if the filter question was omitted but at least one valid response was found in the dependent questions then IEA Hamburg recoded the filter question to "Yes." This of course is only possible for dichotomous filter questions (e.g., with response options such "Yes/No").

IEA Hamburg also applied what are known as split variable checks to questions where the answer was coded into several variables. For example, Question 5 in the student questionnaire asked students: "Do you have any of these things at your home?" Student responses were captured in a set of eight variables, each one coded as "Yes" if the corresponding "Yes" option was filled in and "No" if the "No" option was filled in. Occasionally, students checked the "Yes" boxes for some items but left the boxes for the remaining items unchecked. Because, in these cases, it was clear that the unchecked boxes actually meant "No," these responses were recoded accordingly.

In addition, student reports to items on gender and age in the student questionnaire were checked against the tracking information provided by the School Coordinator or Test Administrator during the within-school sampling and test/questionnaire administration process. When information on gender or birth year and month was missing in the student questionnaire but the student participated, this information, when available, was copied over from the tracking data to the questionnaire. If discrepancies were found between existing tracking and questionnaire gender and age data, IEA Hamburg queried the case with the national center, and the national center investigated which source of information was correct. If unresolved, tracking data was trusted over questionnaire data.

Handling of Missing Data

Two types of entries were possible during the PIRLS 2016 data capture: valid data values and missing data values. Missing data can be assigned a value of "omitted/invalid" or "not administered" during data capture. IEA Hamburg applied additional missing codes to the data to facilitate further analyses. This process led to four distinct types of missing data in the international database:

- Omitted or invalid: The respondent had a chance to answer the question but did not do so, leaving the corresponding item or question blank. This code was also used if the response was uninterpretable or out-of-range.
- Not administered: This signified that the item or question was not administered to the respondent, which meant that the respondent could not read and answer the question. The not administered missing code was used for those student test items that were not in the set of assessment blocks administered to a student either deliberately (due to the rotation of assessment blocks) or in rare cases due to technical failure or incorrect translations. This missing code was also used for those records that were included in the international database but did not contain a single response to one of the assigned questionnaires. For example, this situation applied to home questionnaire data for students who participated in the student test but the parent/guardian did not answer the home questionnaire. In addition, the not administered code was used for individual questionnaire items that a national center decided not to include in the country-specific version of the questionnaire.
- Logically not applicable: The respondent answered a preceding filter question in a way that made the following dependent questions not relevant to him or her.
- Not reached: This applied only to the individual items of the student achievement test and indicated those items that students did not attempt due to a lack of time. "Not reached" codes were derived as follows: First, the last answer given by a student in a session is identified. This could be either a valid or invalid response to an item. The first omitted response after this last answer is coded as "omitted," but all following responses to these items in the session are then coded as "not reached." For example, the response pattern "1 9 4 2 9 9 9 9 9 9" (where "9" represents "omitted") is recoded to "1 9 4 2 9 R R R R R" (where "R" represents "not reached").

Data Cleaning Quality Control

Because PIRLS 2016 was a large and highly complex study with very high standards for data quality, maintaining these standards required an extensive set of interrelated data checking and data cleaning procedures. To ensure that all procedures were conducted in the correct sequence, that no special requirements were overlooked, and that the cleaning process was implemented independently of the persons in charge, the data quality control process included the following steps:

- Thorough testing of all data cleaning programs: Before applying the programs to real datasets, IEA Hamburg applied them to simulation datasets containing all possible problems and inconsistencies
- Registering all incoming data and documents in a specific database: IEA Hamburg recorded the date of arrival as well as specific issues requiring attention
- Carrying out data cleaning according to strict rules: Deviations from the cleaning sequence were not possible, and the scope for involuntary changes to the cleaning procedures was minimal
- Documenting all systematic data recodings that applied to all countries: IEA Hamburg recorded all changes to data in the comprehensive cleaning documentation provided to national centers
- Logging every "manual" correction to a country's data files in a recoding script: Logging these changes, which occurred only occasionally, allowed IEA staff to undo changes or to redo the whole manual cleaning process at any later stage of the data cleaning process
- Repeating, on completion of data cleaning for a country, all cleaning steps from the beginning: This step allowed IEA Hamburg to detect any problems that might have been inadvertently introduced during the data cleaning process
- Working closely with national centers at various steps of the cleaning process: IEA Hamburg provided national centers with the processed data files and accompanying documentation so that center staff could thoroughly review and correct any identified inconsistencies

IEA Hamburg compared national adaptations recorded in the documentation for the national datasets with the structure of the submitted national data files. IEA staff then recorded any identified deviations from the international data structure in the national adaptation database and for the supplementary materials provided with the *PIRLS 2016 User Guide for the International Database*. Whenever possible, IEA Hamburg recoded national deviations to ensure consistency with the international data structure.

Interim Data Products

Before the PIRLS 2016 International Databases were finalized, two major interim versions of the data files were sent to each country—each country receiving only its own data. The first version was sent as soon as the data could be considered "clean" as regards identification codes and linkage issues. Documentation, with a list of the cleaning checks and corrections made in the data, was included to enable the NRC to review the cleaning process before the 7th NRC meeting in Agadir, Morocco in December 2016. A second version of the data files was sent to countries when the weights and international achievement scores were available and had been merged with the data files. This version, sent to countries in advance of the 8th NRC meeting in Riga, Latvia in June 2017, contained only records that satisfied the sampling standards, and allowed the NRCs to replicate the results presented in the international reports.

Interim data products were accompanied by detailed data processing and national adaptation documentation, codebooks, and summary statistics. The summary statistics, preliminary versions of the PIRLS 2016 Almanacs, were created by the TIMSS & PIRLS International Study Center and included weighted univariate statistics for all questionnaire variables for each country. For categorical variables, representing the majority of variables, the percentages of respondents choosing each of the response options were displayed. For continuous numeric variables, various descriptive statistics were reported, including the minimum, maximum, mean, standard deviation, median, mode, and percentiles. For both types of variables, the percentages of missing data were reported. Additionally, for the achievement items, the TIMSS & PIRLS International Study Center provided item analysis and reliability statistics listing information regarding the number of valid cases, percentages, percent correct, Rasch item difficulty, scoring reliability, and so forth. These statistics were used for a more in-depth review of the data at the international and national levels in terms of plausibility, unexpected response patterns, etc. More information on reviewing item statistics is available in Chapter 10.

Final Product – the PIRLS 2016 International Databases

The data cleaning effort implemented at IEA Hamburg ensured that the PIRLS 2016 international databases contained high-quality data. More specifically, the process ensured that:

- Information coded in each variable was internationally comparable
- National adaptations were reflected appropriately in all variables
- All entries in the database could be successfully linked within and across levels
- Sampling weights and student achievement scores were available for international comparisons

Supplements to the *PIRLS 2016 User Guide for the International Database* document all national adaptations made to questionnaires by individual countries and how they were handled in the data. The meaning of country-specific items also can be found in this supplement, as well as recoding requirements by the TIMSS & PIRLS International Study Center.

PIRLS
2016

Methods and Procedures in PIRLS 2016

Reporting

CHAPTER 10

Reviewing the PIRLS 2016 Achievement Item Statistics

Pierre Foy
Michael O. Martin
Ina V.S. Mullis
Liqun Yin

The TIMSS & PIRLS International Study Center conducted an in depth review of a range of diagnostic statistics to examine and evaluate the psychometric characteristics of each achievement item across the countries that participated in the PIRLS 2016 assessments. This review of item statistics is essential to the successful application of item response theory (IRT) scaling to derive student achievement scores for analysis and reporting. The review played a crucial role in the quality assurance of the PIRLS 2016 achievement data prior to scaling, making it possible to detect unusual item properties that could signal a problem or error for a particular country. For example, an item that was uncharacteristically easy or difficult, or had an unusually low discriminating power, could indicate a potential problem with either translation or printing. Similarly, a constructed response item with unusually low scoring reliability could indicate a problem with a scoring guide in a particular country. In the rare instances where such items were found, the country's translation verification documents and printed booklets were examined for flaws or inaccuracies and, if necessary, the item was removed from the international database for that country.

Statistics for Item Review

The TIMSS & PIRLS International Study Center computed item statistics for all achievement items in the 2016 assessments, including PIRLS (175 items), PIRLS Literacy (183 items), and ePIRLS (91 items). The item statistics for each of the participating countries were then carefully reviewed. Exhibits 10.1 and 10.2 show actual samples of the statistics calculated for a multiple-choice and a constructed response item, respectively.

Exhibit 10.1: Example International Item Statistics for a PIRLS 2016 Multiple-Choice Item

Progress in International Reading Literacy Study - PIRLS 2016 Assessment Results
International Item Review Statistics (Unweighted)

Acquire and Use Information - The Green Sea Turtle (R41T09M) Feeding habits of an adult green sea turtle
Make Straightforward Inferences - 1 Point - Key: C

Country	Cases	DIFF	DISC	Percentages P_A	P_B	P_C	P_D	P_OM	P_NR	Point Biserials PB_A	PB_B	PB_C	PB_D	PB_OM	PB_NR	RDIFF	Avg. Score Girls	Boys	Flags
Australia	1043	57.0	0.52	11.0	17.5	57.0	11.5	2.9	1.8	-0.21	-0.25	0.52	-0.23	-0.15	-0.18	0.78	59.1	54.7	E
Austria	731	53.3	0.53	14.3	14.8	53.3	14.7	2.9	1.4	-0.17	-0.25	0.53	-0.27	-0.13	-0.14	0.82	52.6	53.8	
Azerbaijan	993	32.2	0.32	24.8	24.9	32.2	14.6	3.4	2.3	-0.10	-0.06	0.32	-0.21	-0.02	-0.08	1.14	32.5	31.9	H
Bahrain	905	29.4	0.31	22.7	24.2	29.4	18.9	4.8	3.5	-0.09	-0.14	0.31	-0.07	-0.06	-0.00	0.96	28.8	30.1	
Belgium (Flemish)	855	53.5	0.44	12.9	17.5	53.5	13.0	3.2	1.8	-0.17	-0.18	0.44	-0.23	-0.09	-0.08	0.71	55.1	51.6	E
Belgium (French)	773	46.4	0.47	16.0	15.9	46.4	16.3	5.3	3.2	-0.19	-0.19	0.47	-0.18	-0.12	-0.06	0.61	48.1	44.6	E
*Bulgaria	704	71.7	0.48	7.3	11.1	71.7	9.4	0.6	0.3	-0.22	-0.27	0.48	-0.25	-0.04	-0.08	0.16	72.5	70.8	E F
*Canada	2980	54.0	0.48	12.9	15.5	54.0	13.4	4.3	4.1	-0.18	-0.23	0.48	-0.20	-0.14	-0.17	0.82	55.8	52.2	E
Chile	715	30.8	0.41	19.8	26.4	30.8	19.7	3.2	4.8	-0.17	-0.14	0.41	-0.13	-0.06	-0.06	1.49	29.6	32.0	H
Chinese Taipei	719	46.6	0.40	19.6	21.5	46.6	11.5	0.8	0.4	-0.22	-0.08	0.40	-0.23	-0.03	-0.02	1.44	47.8	45.5	H
*Czech Republic	906	53.5	0.53	15.3	14.3	53.5	12.4	4.7	0.9	-0.25	-0.18	0.53	-0.27	-0.13	-0.05	1.07	56.7	50.3	H
Denmark	557	59.2	0.57	14.8	10.1	59.2	12.6	3.3	2.0	-0.17	-0.32	0.57	-0.31	-0.11	-0.12	0.67	62.3	55.8	E
*England	846	56.7	0.55	15.4	14.1	56.7	11.1	2.7	0.8	-0.25	-0.23	0.55	-0.25	-0.13	-0.07	0.99	57.3	56.1	
Finland	811	59.1	0.56	9.0	15.3	59.1	13.6	3.0	0.2	-0.21	-0.25	0.56	-0.30	-0.16	-0.10	1.05	62.8	55.5	H F G
*France	783	48.8	0.50	15.6	17.2	48.8	13.4	5.1	4.0	-0.21	-0.22	0.50	-0.19	-0.12	-0.10	0.56	49.7	47.8	E
Georgia	924	43.4	0.45	16.0	23.5	43.4	13.1	4.0	4.2	-0.18	-0.18	0.45	-0.17	-0.11	-0.10	0.48	45.5	41.1	E
*Germany	654	62.2	0.55	10.1	9.8	62.2	13.2	4.8	1.4	-0.27	-0.24	0.55	-0.30	-0.07	-0.12	0.54	62.2	62.2	E F
*Hong Kong SAR	561	53.2	0.39	18.2	16.8	53.2	10.2	1.6	0.2	-0.12	-0.23	0.39	-0.13	-0.18	0.00	1.30	55.2	51.2	H
*Hungary	762	65.8	0.54	12.9	11.5	65.8	7.9	1.8	0.5	-0.27	-0.26	0.54	-0.25	-0.13	-0.02	0.59	69.2	62.4	E F G
*Iran, Islamic Rep. of	725	26.7	0.20	25.3	16.6	26.7	27.3	4.2	3.9	0.02	-0.05	0.20	-0.13	-0.08	-0.05	0.74	25.7	27.7	E
Ireland	769	63.4	0.55	11.8	12.3	63.4	10.7	1.7	0.8	-0.21	-0.27	0.55	-0.28	-0.17	-0.11	0.70	63.3	63.5	E
*Israel	674	37.8	0.44	17.1	23.3	37.8	17.1	4.7	6.2	-0.19	-0.17	0.44	-0.09	-0.16	-0.18	1.49	38.5	37.1	H
*Italy	653	54.5	0.50	10.8	15.0	54.5	17.5	2.2	1.1	-0.20	-0.31	0.50	-0.16	-0.12	-0.13	1.04	57.8	51.5	H
Kazakhstan	818	40.6	0.47	12.0	23.4	40.6	22.3	1.7	0.2	-0.11	-0.23	0.47	-0.21	-0.07	-0.03	1.38	42.8	38.7	H
*Latvia	678	58.8	0.52	11.7	15.9	58.8	11.3	2.4	0.4	-0.23	-0.22	0.52	-0.28	-0.10	-0.08	0.95	60.9	56.4	
*Lithuania	709	56.9	0.53	13.5	12.5	56.9	14.8	2.4	0.6	-0.25	-0.18	0.53	-0.27	-0.15	-0.07	0.70	58.7	55.1	E
Macao SAR	678	57.8	0.49	18.4	12.8	57.8	9.8	1.2	0.7	-0.29	-0.20	0.49	-0.19	-0.05	-0.02	0.73	55.6	59.7	E F
Malta	593	25.2	0.32	24.4	24.4	25.2	20.8	5.3	1.0	-0.08	-0.15	0.32	-0.05	-0.11	-0.03	0.82	29.4	21.0	G
*Morocco	904	25.3	0.22	23.8	25.6	25.3	18.8	6.5	6.3	-0.05	-0.07	0.22	-0.03	-0.15	-0.03	0.01	23.2	27.3	E
*Netherlands	682	59.6	0.56	11.6	13.7	59.6	13.1	2.1	1.3	-0.22	-0.27	0.56	-0.32	-0.05	-0.05	0.63	61.4	57.9	E
*New Zealand	909	54.9	0.49	13.9	14.2	54.9	13.8	3.1	2.2	-0.22	-0.23	0.49	-0.19	-0.13	-0.13	0.53	54.2	55.6	E
Northern Ireland	602	58.3	0.43	10.5	14.2	58.3	14.4	2.7	0.5	-0.25	-0.25	0.43	-0.13	-0.03	-0.01	1.08	58.2	58.3	H
*Norway	682	58.8	0.52	11.7	12.8	58.8	14.3	2.4	1.3	-0.15	-0.26	0.52	-0.30	-0.09	-0.11	0.94	62.6	55.1	G
Oman	1524	26.5	0.20	22.8	23.4	26.5	21.0	6.3	3.5	-0.00	-0.10	0.20	-0.04	-0.10	-0.07	0.84	25.0	28.1	
Poland	720	45.0	0.53	18.0	14.9	45.0	20.1	2.0	0.4	-0.20	-0.24	0.53	-0.23	-0.09	-0.01	1.67	48.0	41.7	H
Portugal	768	38.0	0.43	17.3	24.8	38.0	15.8	4.0	3.0	-0.17	-0.20	0.43	-0.13	-0.07	-0.07	1.34	38.1	37.9	H
Qatar	1475	30.2	0.38	21.4	25.9	30.2	17.6	5.0	3.9	-0.05	-0.18	0.38	-0.14	-0.10	-0.09	0.98	30.9	29.4	
*Russian Federation	770	63.8	0.52	8.6	11.1	63.8	14.1	2.3	0.5	-0.19	-0.29	0.52	-0.23	-0.16	-0.08	1.08	67.1	60.4	H F
Saudi Arabia	783	27.3	0.30	22.6	26.9	27.3	17.7	5.6	6.0	-0.18	-0.11	0.30	0.02	-0.08	-0.09	0.92	30.6	23.8	G
*Singapore	1075	63.0	0.50	10.8	15.7	63.0	9.0	1.5	0.4	-0.24	-0.28	0.50	-0.21	-0.07	-0.11	0.84	65.1	60.7	F
*Slovak Republic	904	52.6	0.49	14.1	14.5	52.6	15.9	2.9	0.7	-0.19	-0.14	0.49	-0.31	-0.10	-0.09	0.85	52.0	53.0	
*Slovenia	748	55.5	0.50	9.9	13.7	55.5	17.5	3.4	1.5	-0.19	-0.25	0.50	-0.24	-0.06	-0.10	0.78	52.8	58.0	F
Spain	2408	41.7	0.47	17.6	16.3	41.7	22.5	1.9	0.9	-0.15	-0.14	0.47	-0.28	-0.06	-0.04	1.38	41.8	41.5	H
*Sweden	751	61.7	0.53	12.4	10.7	61.7	12.2	3.0	2.0	-0.22	-0.23	0.53	-0.30	-0.10	-0.18	0.80	63.5	59.9	
Trinidad and Tobago	682	37.8	0.49	18.8	22.0	37.8	16.7	4.7	5.4	-0.16	-0.25	0.49	-0.13	-0.11	-0.08	0.80	37.7	38.0	
United Arab Emirates	2730	37.7	0.43	18.0	22.6	37.7	16.4	5.3	2.1	-0.13	-0.20	0.43	-0.11	-0.15	-0.09	0.83	39.8	35.9	G
*United States	711	49.1	0.50	14.3	18.4	49.1	15.3	2.9	1.5	-0.23	-0.25	0.50	-0.17	-0.09	-0.11	1.13	48.3	50.2	H
*Reference Avg. (23)	19771	54.1	0.48	13.8	14.9	54.1	14.0	3.1	1.8	-0.20	-0.22	0.48	-0.22	-0.11	-0.09	0.80	55.2	53.0	G
International Avg. (47)	43347	48.6	0.46	15.6	17.4	48.6	15.1	3.3	2.0	-0.18	-0.20	0.46	-0.20	-0.10	-0.08	0.90	49.7	47.5	G
Buenos Aires, Argentina	704	34.3	0.47	20.4	17.1	34.3	20.1	8.0	11.1	-0.18	-0.17	0.47	-0.13	-0.11	-0.09	0.99	35.6	33.1	
Ontario, Canada	692	50.7	0.45	11.5	18.4	50.7	14.4	5.0	4.8	-0.13	-0.26	0.45	-0.13	-0.16	-0.16	0.96	54.1	47.3	
Quebec, Canada	523	58.6	0.43	13.1	14.7	58.6	10.8	2.8	4.0	-0.15	-0.20	0.43	-0.25	-0.08	-0.17	0.89	59.8	57.2	
Norway (4)	692	49.6	0.50	19.7	13.6	49.6	14.4	2.7	4.5	-0.21	-0.19	0.50	-0.23	-0.12	-0.13	0.74	50.1	49.1	E
Moscow City, Russian Fed.	718	66.2	0.52	8.8	8.5	66.2	14.9	1.5	0.1	-0.24	-0.26	0.52	-0.27	-0.08	-0.03	1.39	71.4	61.3	H F G
Eng/Afr/Zulu - RSA (5)	854	25.4	0.38	21.3	25.5	25.4	23.1	4.6	3.3	-0.13	-0.11	0.38	-0.09	-0.12	-0.05	0.55	28.0	22.5	E
Andalusia, Spain	687	37.7	0.47	17.2	17.2	37.7	26.1	1.8	0.9	-0.12	-0.09	0.47	-0.32	-0.05	-0.06	1.31	35.9	39.6	H
Madrid, Spain	628	47.8	0.48	16.0	15.7	47.8	18.7	1.8	0.5	-0.21	-0.17	0.48	-0.23	-0.11	-0.08	1.35	47.7	48.0	H
Abu Dhabi, UAE	683	32.4	0.35	19.0	23.3	32.4	17.8	7.5	2.8	-0.10	-0.12	0.35	-0.09	-0.14	-0.11	0.64	33.7	31.3	E
Dubai, UAE	1320	44.4	0.46	17.0	20.8	44.4	14.6	3.2	1.3	-0.16	-0.27	0.46	-0.11	-0.12	-0.07	0.97	46.8	42.3	

Keys: DIFF= Percent correct score; DISC= Item discrimination; P_A...P_D= Percentage choosing each option; P_OM, P_NR= Percentage Omitted, Not Reached; PB_A...PB_D= Point Biserial for each option; PB_OM, PB_NR= Point Biserial for Omitted, Not Reached; RDIFF= Rasch difficulty.
Flags: A= Attractive distractor; B= Boys outperform girls; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average; F= Distractor chosen by less than 10%; G= Girls outperform boys; H= Harder than average; R= Scoring reliability less than 85%; V= Difficulty greater than 95%.

Exhibit 10.2: Example International Item Statistics for a PIRLS 2016 Constructed Response Item

Progress in International Reading Literacy Study - PIRLS 2016 Assessment Results
International Item Review Statistics (Unweighted)

Literary Experience - Oliver and The Griffin (R41O13C) Describe why griffin can not do his job
Interpret and Integrate Ideas and Information - 3 Points

				Percentages						Point Biserials							Reliability		Avg. Score		
Country	Cases	DIFF	DISC	P_0	P_1	P_2	P_3	P_OM	P_NR	PB_0	PB_1	PB_2	PB_3	PB_OM	PB_NR	RDIFF	N	Agr	Girls	Boys	Flags
Australia	1047	65.4	0.70	10.4	19.0	19.1	46.4	5.1	3.4	-0.43	-0.24	0.06	0.56	-0.35	-0.18	0.33	224	85.3	70.4	60.9	__E___G
Austria	726	49.6	0.61	29.1	6.5	24.6	31.0	8.8	1.9	-0.41	0.01	0.10	0.50	-0.31	-0.10	1.16	218	97.2	48.6	50.5	__H_F__
Azerbaijan	994	14.8	0.48	53.8	10.3	8.4	5.7	21.8	3.4	-0.13	0.17	0.28	0.29	-0.33	-0.12	1.73	241	97.5	15.9	13.8	__H_F__
Bahrain	911	22.0	0.65	54.8	7.6	10.5	12.5	14.6	7.7	-0.40	0.08	0.29	0.51	-0.22	-0.13	0.88	239	91.6	24.4	19.6	__E_F__
Belgium (Flemish)	860	46.0	0.61	35.6	6.8	19.3	30.8	7.5	2.3	-0.40	-0.06	0.15	0.51	-0.32	-0.10	0.96	256	93.0	48.8	43.0	____F__
Belgium (French)	769	38.3	0.59	33.8	9.1	20.3	21.7	15.1	5.9	-0.33	0.01	0.24	0.43	-0.33	-0.15	0.91	225	97.8	39.1	37.4	__E_F__
*Bulgaria	712	47.2	0.63	32.8	6.2	26.7	27.3	7.0	2.2	-0.41	-0.05	0.13	0.52	-0.33	-0.13	1.44	201	92.0	49.8	44.5	__H_F__
*Canada	2996	56.3	0.65	27.6	8.2	14.3	44.0	5.8	6.3	-0.47	-0.06	0.07	0.56	-0.30	-0.21	0.78	456	80.7	60.4	52.2	__E_FRG
Chile	712	42.3	0.66	38.0	9.7	17.7	27.2	7.3	6.2	-0.47	-0.04	0.20	0.53	-0.28	-0.12	0.84	207	96.1	45.6	39.4	__E_F__
Chinese Taipei	722	48.7	0.64	28.2	4.1	19.0	34.6	14.1	1.7	-0.40	-0.01	0.13	0.52	-0.34	-0.11	1.40	213	97.7	51.2	46.1	__H_F__
*Czech Republic	923	42.2	0.60	31.4	8.2	18.8	26.9	14.6	2.7	-0.29	-0.01	0.18	0.47	-0.41	-0.14	1.38	215	100.0	43.8	40.7	__H_F__
Denmark	593	47.7	0.59	33.3	8.5	22.3	30.0	5.9	4.9	-0.42	-0.08	0.14	0.48	-0.26	-0.18	1.26	182	86.3	50.8	44.4	__H_F__
*England	849	62.6	0.67	24.2	5.7	13.5	51.7	5.0	2.5	-0.49	-0.05	-0.03	0.61	-0.35	-0.17	0.82	245	94.7	65.2	59.7	__E_F__
Finland	807	61.4	0.59	22.6	7.2	15.2	48.8	6.1	2.5	-0.39	-0.09	0.00	0.53	-0.34	-0.17	1.11	229	93.9	63.1	59.8	__H_F__
*France	795	42.8	0.65	33.2	6.4	14.2	31.2	15.0	7.7	-0.40	-0.00	0.12	0.57	-0.32	-0.18	0.82	232	96.1	44.1	41.4	__E_F__
Georgia	954	30.1	0.66	41.4	9.0	14.4	17.5	17.7	8.9	-0.33	0.12	0.23	0.52	-0.40	-0.22	1.19	206	81.6	33.9	26.3	__H_FRG
*Germany	649	49.7	0.67	17.7	20.7	23.0	27.4	11.2	2.3	-0.41	-0.09	0.17	0.51	-0.33	-0.16	1.06	228	90.8	51.3	48.1	_______
*Hong Kong SAR	552	56.4	0.59	19.9	9.6	18.4	40.9	11.2	1.6	-0.30	-0.17	0.11	0.49	-0.36	0.01	1.25	170	92.4	57.6	55.3	__H_F__
*Hungary	775	57.1	0.62	21.3	16.8	19.9	38.2	3.8	1.4	-0.43	-0.11	0.04	0.52	-0.26	-0.15	0.96	210	94.3	57.1	57.0	_______
*Iran, Islamic Rep. of	735	31.4	0.70	35.2	14.5	13.2	17.8	19.3	13.5	-0.40	0.08	0.28	0.53	-0.33	-0.23	0.76	215	92.1	34.1	28.5	__E____
Ireland	762	66.0	0.60	21.8	4.3	18.2	52.4	3.3	1.8	-0.43	-0.10	-0.03	0.53	-0.32	-0.09	0.92	209	98.1	68.0	64.1	__E_F__
*Israel	660	45.9	0.69	32.1	5.5	16.2	33.2	13.0	9.2	-0.46	0.01	0.22	0.54	-0.36	-0.23	1.14	178	98.3	47.3	44.3	__H_F__
*Italy	661	62.9	0.58	20.5	5.1	16.4	50.2	7.8	2.4	-0.33	-0.06	-0.02	0.52	-0.40	-0.22	0.85	188	94.1	65.9	60.0	__E_F__
Kazakhstan	818	51.1	0.64	19.0	24.8	28.3	24.0	4.0	1.3	-0.45	-0.13	0.16	0.47	-0.20	-0.15	0.85	215	97.2	53.7	48.6	__E___G
*Latvia	702	49.2	0.55	33.4	10.2	18.5	33.4	4.4	0.7	-0.37	-0.12	0.13	0.46	-0.28	-0.08	1.36	214	94.4	52.6	45.8	__H___G
*Lithuania	725	41.1	0.58	38.5	7.3	19.2	25.8	9.1	1.8	-0.32	-0.05	0.11	0.51	-0.34	-0.12	1.48	220	99.1	40.0	42.3	__H_F__
Macao SAR	673	43.7	0.60	31.5	5.3	17.0	30.6	15.5	2.4	-0.32	-0.01	0.15	0.49	-0.37	-0.11	1.33	219	99.1	43.1	44.4	__H_F__
Malta	604	27.1	0.56	50.4	6.5	11.7	17.2	14.2	5.5	-0.25	0.02	0.14	0.51	-0.33	-0.16	0.74	191	85.3	29.4	24.8	__E_F__
*Morocco	909	11.5	0.65	55.8	7.7	5.1	5.6	25.8	13.1	-0.31	0.24	0.28	0.50	-0.19	-0.06	0.36	217	84.8	13.8	9.2	__E_FRG
*Netherlands	705	52.0	0.58	34.9	7.0	18.1	37.6	2.3	3.0	-0.47	-0.09	0.05	0.52	-0.17	-0.14	1.11	161	96.9	53.6	50.4	__H_F__
*New Zealand	942	56.2	0.68	31.3	5.5	13.5	45.4	4.4	5.0	-0.52	-0.06	0.09	0.58	-0.32	-0.24	0.72	248	94.8	61.0	51.1	__E_F_G
Northern Ireland	608	58.4	0.58	28.0	5.2	15.9	46.1	4.9	1.8	-0.39	-0.10	-0.02	0.54	-0.29	-0.18	1.02	174	100.0	60.2	56.7	____F__
*Norway	719	59.3	0.61	22.9	7.5	21.9	42.2	5.5	2.1	-0.38	-0.14	0.02	0.54	-0.34	-0.17	1.13	213	96.7	62.4	56.4	__H_F__
Oman	1546	16.6	0.66	57.3	11.7	8.2	7.2	15.5	8.5	-0.45	0.21	0.35	0.45	-0.15	-0.13	0.97	210	91.9	20.3	13.0	____F_G
Poland	725	60.2	0.69	14.1	13.8	33.8	33.1	5.2	1.2	-0.48	-0.24	0.15	0.49	-0.25	-0.22	1.12	183	83.1	63.0	57.8	_____R_
Portugal	763	43.0	0.59	38.3	3.4	19.4	28.9	10.0	8.3	-0.39	-0.04	0.12	0.51	-0.27	-0.16	1.14	235	98.7	44.2	41.6	__H_F__
Qatar	1497	26.6	0.72	50.6	8.8	12.1	15.6	12.9	6.9	-0.52	0.05	0.29	0.58	-0.17	-0.12	0.76	224	97.3	29.7	23.2	__E_F_G
*Russian Federation	771	44.1	0.54	42.4	6.1	17.1	30.7	3.8	1.4	-0.39	-0.08	0.07	0.49	-0.22	-0.09	1.98	206	97.1	46.3	41.8	__H_F__
Saudi Arabia	766	23.3	0.63	39.3	21.3	12.8	7.7	18.9	10.1	-0.31	0.16	0.32	0.40	-0.33	-0.11	0.72	199	98.5	28.8	17.5	__E_F_G
*Singapore	1081	68.0	0.68	22.1	2.1	16.6	56.2	2.9	0.9	-0.57	-0.08	0.00	0.58	-0.23	-0.10	0.89	239	100.0	70.4	65.9	__E_F__
*Slovak Republic	910	46.1	0.59	37.6	5.4	18.5	32.0	6.4	1.0	-0.40	0.01	0.14	0.47	-0.34	-0.07	1.23	219	95.9	45.1	47.1	__H_F__
*Slovenia	748	49.6	0.64	31.3	6.4	19.9	34.2	8.3	3.3	-0.41	-0.06	0.12	0.53	-0.34	-0.19	1.11	225	99.6	50.6	48.4	____F__
Spain	2427	52.4	0.60	30.7	4.7	23.4	35.2	6.0	2.6	-0.41	-0.06	0.13	0.48	-0.32	-0.14	0.92	402	98.3	52.7	52.0	__E_F__
*Sweden	748	58.8	0.60	25.6	4.2	15.5	47.1	7.7	4.3	-0.39	-0.08	0.06	0.52	-0.35	-0.19	1.10	182	94.5	59.2	58.3	____F__
Trinidad and Tobago	702	36.3	0.67	43.5	6.9	14.0	24.6	11.0	9.3	-0.48	0.02	0.16	0.58	-0.24	-0.23	0.85	196	89.3	37.7	34.7	__E_F__
United Arab Emirates	2750	30.0	0.71	51.0	5.7	11.4	20.5	11.3	5.8	-0.52	0.08	0.23	0.59	-0.22	-0.16	0.85	687	88.4	32.0	28.1	__E_F_G
*United States	745	63.0	0.65	24.6	6.6	16.3	49.9	2.5	4.7	-0.53	-0.12	0.03	0.56	-0.23	-0.24	0.76	197	94.4	65.8	60.3	__E_F__
*Reference Avg. (23)	20012	50.1	0.63	30.3	7.9	17.2	36.0	8.6	4.1	-0.41	-0.05	0.10	0.53	-0.31	-0.15	1.07	5079	94.5	52.1	48.2	____F_G
International Avg. (47)	43748	45.8	0.63	33.0	8.6	17.3	31.5	9.7	4.4	-0.40	-0.03	0.14	0.51	-0.30	-0.15	1.03	10863	94.0	47.9	43.8	____F_G
Buenos Aires, Argentina	720	34.3	0.60	39.0	7.5	8.9	25.9	18.7	11.4	-0.32	0.01	0.13	0.54	-0.31	-0.26	0.84	200	95.0	33.6	35.0	__E_F__
Ontario, Canada	708	56.4	0.69	27.3	9.2	14.3	43.9	5.4	7.9	-0.52	-0.06	0.04	0.60	-0.29	-0.25	0.64	108	81.5	61.3	51.5	__E_FRG
Quebec, Canada	526	58.4	0.61	25.1	7.5	15.0	45.8	6.5	3.8	-0.42	-0.09	0.07	0.53	-0.32	-0.08	1.14	103	84.5	62.3	54.3	__H_FRG
Norway (4)	734	47.4	0.59	24.3	9.6	25.4	27.3	13.4	7.6	-0.34	-0.04	0.15	0.46	-0.33	-0.26	0.98	.	.	51.0	43.7	____F_G
Moscow City, Russian Fed.	714	58.0	0.54	29.6	3.4	22.8	41.6	2.5	0.7	-0.38	-0.14	-0.02	0.50	-0.26	-0.09	1.87	219	97.7	60.7	55.2	__H_F__
Eng/Afr/Zulu - RSA (5)	874	16.6	0.68	67.5	12.6	7.4	7.4	5.0	9.2	-0.52	0.17	0.32	0.51	-0.15	-0.09	0.79	208	84.6	19.3	13.7	__E_FRG
Andalusia, Spain	688	47.0	0.60	35.3	3.0	24.5	29.6	7.5	3.3	-0.39	-0.04	0.18	0.46	-0.36	-0.16	1.08	200	98.0	47.1	46.8	____F__
Madrid, Spain	636	57.9	0.62	28.5	2.4	24.6	40.6	3.8	1.7	-0.50	0.02	0.11	0.47	-0.30	-0.13	1.01	86	98.8	58.2	57.6	____F__
Abu Dhabi, UAE	701	20.4	0.67	55.7	6.2	8.3	12.8	17.0	7.6	-0.41	0.13	0.22	0.57	-0.20	-0.13	0.82	211	88.2	24.3	16.5	__E_F_G
Dubai, UAE	1311	42.5	0.70	41.0	5.5	15.0	30.7	7.8	3.5	-0.54	0.02	0.17	0.59	-0.25	-0.18	0.87	259	88.0	43.6	41.5	__E_F__

Keys: DIFF= Percent correct score; DISC= Item discrimination; P_0...P_3= Percentage obtaining score level; P_OM, P_NR= Percentage Omitted, Not Reached;
PB_0...PB_3= Point Biserial for score level; PB_OM, PB_NR= Point Biserial for Omitted, Not Reached; RDIFF= Rasch difficulty;
Reliability: N= Responses double scored; Agr= Percentage agreement.
Flags: A= Point-biserial not ordered; B= Boys outperform girls; C= Difficulty less than chance; D= Negative/low discrimination; E= Easier than average;

For all items, regardless of format (i.e., multiple-choice or constructed response), statistics included the number of students that responded in each country, the difficulty level (the percentage of students that answered the item correctly), and the discrimination index (the point-biserial correlation between success on the item and total score).[1] Also provided was an estimate of the difficulty of the item using a Rasch one-parameter IRT model. Statistics for each item were displayed alphabetically by country, together with an international average—i.e., based on all participating countries listed above the international average—and a reference average—based on a pool of countries that have participated regularly in the PIRLS assessments—for each statistic. The reference countries are shown with an asterisk next to their names. The international and reference averages of the item difficulties and item discriminations served as guides to the overall statistical properties of the items. The item review outputs also listed the benchmarking participants.

Statistics displayed for multiple-choice items included the percentage of students that chose each response option—as well as the percentage of students that omitted or did not reach the item—and the point-biserial correlations for each response option. Statistics displayed for constructed response items (which could have 1, 2, or 3 score points) included the percent correct and point-biserial of each score level. Constructed response item tables also provided information about the reliability with which each item was scored in each country, showing the total number of double-scored responses and the percentage of score agreement between the scorers.

During item review, “not reached” responses (i.e., items toward the end of the booklet that the student did not attempt)[2] were treated as “not administered” and thus did not contribute to the calculation of the item statistics. However, the percentage of students not reaching each item was reported. Omitted responses, although treated as incorrect, were tabulated separately from incorrect responses for the sake of distinguishing students who provided no form of response from students who attempted a response.

The definitions and detailed descriptions of the statistics that were calculated are given below. The statistics are listed in order of their appearance in the item review outputs:

CASES: This is the number of students to whom the item was administered. Not-reached responses were not included in this count.

DIFF: The item difficulty is the average percent correct on an item. For a 1-point item, including all multiple-choice items, it is the percentage of students providing a fully correct response to the item. For 2-point and 3-point items, it is the average percentage of points. For example, if 25 percent of students scored 2 points, 50 percent scored 1 point on a 2-point item, and the other 25 percent score 0 points, then the average percent correct for such an item would be 50 percent. For this statistic, not reached responses were not included.

1 For computing point-biserial correlations, the total score is the percentage of points a student has scored on the items (s)he was administered. Not reached responses are not included in the total score.

2 An item was considered “not reached” if the item itself and the item immediately preceding it were not answered and no subsequent items had been attempted. The decision as to whether an item was not reached was made separately for part 1 and part 2 of each assessment booklet.

DISC: The item discrimination is computed as the correlation between the response to an item and the total score on all items administered to a student. Items exhibiting good measurement properties should have a moderately positive correlation, indicating that the more able students get the item right, the less able get it wrong. For this statistic, not reached items were not included.

PCT_A, PCT_B, PCT_C, and PCT_D: Available for multiple-choice items. Each column indicates the percentage of students choosing the particular response option for the item (A, B, C, or D).[3] Not reached responses were excluded from the denominator.

PCT_0, PCT_1, PCT_2, and PCT_3: Available for constructed response items. Each column indicates the percentage of students responding at that particular score level, up to and including the maximum score level for the item. Not reached items were excluded from the denominator.

PCT_OM: Percentage of students who, having reached the item, did not provide a response. Not reached responses were excluded from the denominator.

PCT_NR: Percentage of students who did not reach the item. This statistic is the number of students who did not reach an item as a percentage of all students who were administered that item, including those who omitted or did not reach that item.

PB_A, PB_B, PB_C, and PB_D: Available for multiple-choice items. These columns show the point-biserial correlations between choosing each of the response options (A, B, C, or D) and the total score on all of the items administered to a student. Items with good psychometric properties have moderately positive correlations for the correct option and negative correlations for the distracters (the incorrect options). Not reached responses were not included in these calculations.

PB_0, PB_1, PB_2, and PB_3: Available for constructed response items. These columns present the point-biserial correlations between the score levels on the item and the overall score on all of the items the student was administered. For items with good measurement properties, the correlation coefficients should monotonically increase from negative to positive as the score on the item increases. Not reached responses were not included in these calculations.

PB_OM: The point-biserial correlation between a binary variable indicating an omitted response to the item, and the total score on all items administered to a student. This correlation should be negative or near zero. Not reached responses were not included in this statistic.

3 ePIRLS included multiple-choice items with as many as six response options, thus adding options E and F.

PB_NR: The point-biserial correlation between a binary variable indicating a not-reached response to the item, and the total score on all items administered to a student. This correlation should be negative or near zero.

RDIFF: An estimate of the difficulty of an item based on a Rasch one-parameter IRT model applied to the achievement data of a given country. The difficulty estimate is expressed in the logit metric (with a positive logit indicating a difficult item) and was scaled so that the average Rasch item difficulty across all items within each country was zero.

Reliability (N): To provide a measure of the reliability of the scoring of the constructed response items, items in approximately 25 percent of the test booklets in each country were independently scored by two scorers. This column indicates the number of responses that were double-scored for a given item in a country.

Reliability (Agr): This column contains the percentage of agreement on the scores assigned by the two independent PIRLS scorers.

As an aid to the reviewers, the item review displays included a series of flags signaling the presence of one or more conditions that might indicate a problem with an item. The following conditions were flagged:

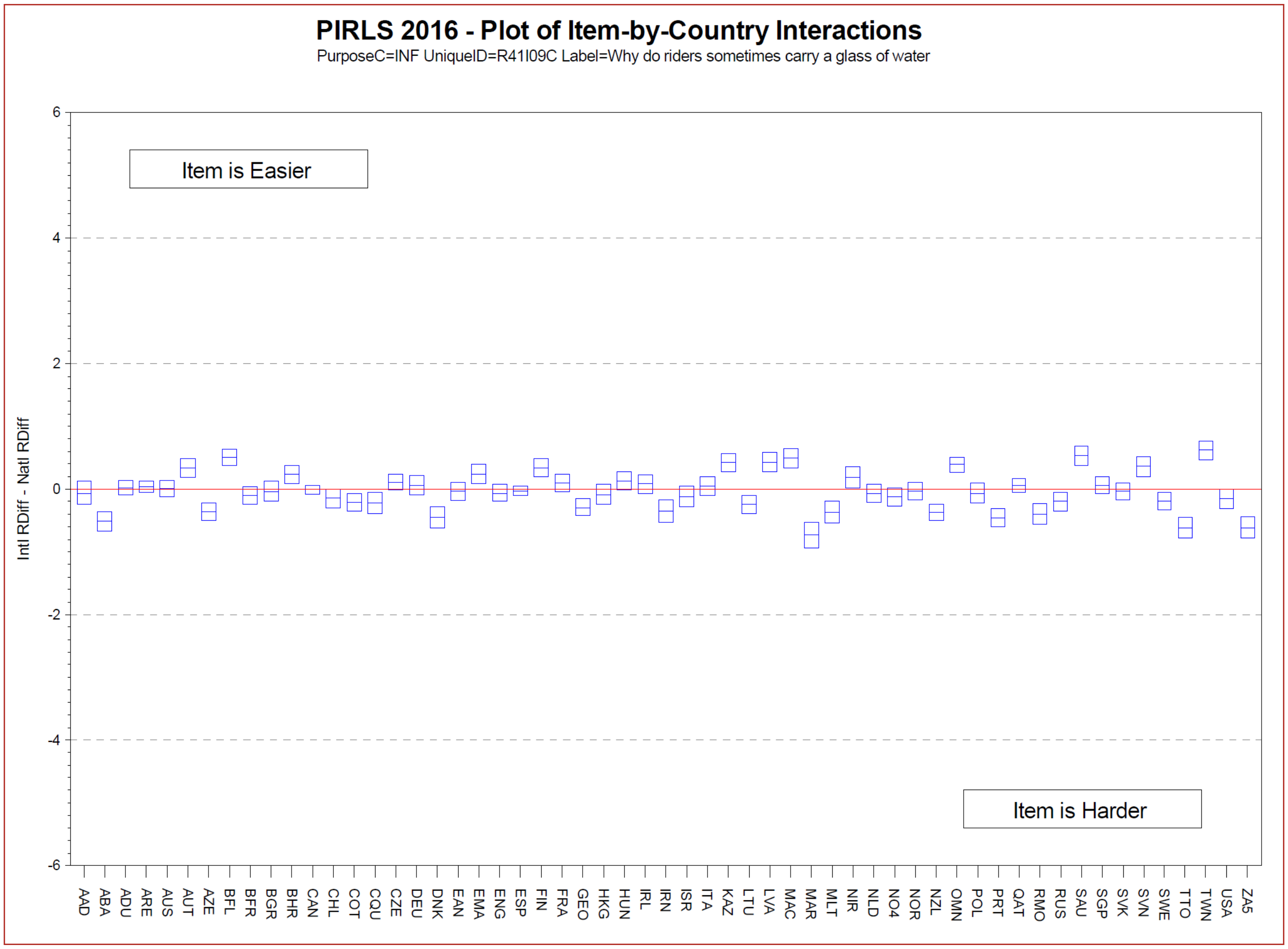
- The item discrimination (DISC) was less than 0.10 (flag D)
- The item difficulty (DIFF) was less than 25% for multiple-choice items (flag C)
- The item difficulty (DIFF) exceeded 95% (flag V)
- The Rasch difficulty estimate (RDIFF) for a given country made the item either easier (flag E) or more difficult (flag H) relative to the international average for that item
- The point-biserial correlation for at least one distracter in a multiple-choice item was positive, or the point-biserial correlations across the score levels of a constructed response item were not ordered (flag A)
- The percentage of students selecting one of the response options for a multiple-choice item, or one of the score values for a constructed response item, was less than 10% (flag F)
- Scoring reliability for agreement on the score value of a constructed response item was less than 85% (flag R)

Although not all of these conditions necessarily indicated a problem, the flags were a useful tool to draw attention to potential sources of concern.

Item–by–Country Interaction

Although countries are expected to exhibit some variation in performance across items, in general countries with high average performance on the assessment should perform relatively well on each of the items, and low-scoring countries should do less well on each of the items. When this does not occur (e.g., when a high-performing country has low performance on an item on which other countries are doing well), there is said to be an item-by-country interaction. When large, such item-by-country interactions may be a sign that an item is flawed in some way and that steps should be taken to address the problem. To assist in detecting sizeable item-by-country interactions, the TIMSS & PIRLS International Study Center produced a graphical display for each item showing the difference between each country's Rasch item difficulty and the international average Rasch item difficulty across all countries. An example of the graphical displays is provided in Exhibit 10.3.

Exhibit 10.3: Example Plot of Item-by-Country Interaction for a PIRLS 2016 Item



In each of these item-by-country interaction displays, the difference in Rasch item difficulty for each country is presented as a 95 percent confidence interval, which includes a built-in Bonferroni correction for multiple comparisons across the participating countries. The limits for this confidence interval were computed as follows:

$$\text{Upper Limit} = RDIFF_{i.} - RDIFF_{ik} + SE(RDIFF_{ik}) \cdot Z_b \qquad (10.1)$$

$$\text{Lower Limit} = RDIFF_{i.} - RDIFF_{ik} - SE(RDIFF_{ik}) \cdot Z_b \qquad (10.2)$$

where $RDIFF_{ik}$ is the Rasch difficulty of item i in country k, $RDIFF_{i.}$ is the international average Rasch difficulty of item i, $SE(RDIFF_{ik})$ is the standard error of the Rasch difficulty of item i in country k, and Z_b is the 95% critical value from the Z distribution corrected for multiple comparisons using the Bonferroni procedure.

Trend Item Review

In order to measure trends, PIRLS 2016 included achievement items from previous assessments as well as items developed for use for the first time in 2016. Accordingly, the PIRLS 2016 assessments included items from 2001, 2006, 2011, and 2016. An important review step, therefore, was to check that these "trend items" had statistical properties in 2016 similar to those they had in the previous assessments (e.g., a PIRLS item that was relatively easy in 2011 should still be relatively easy in 2016).

As can be seen in the example in Exhibit 10.4, the trend item review focused on statistics for trend items from the current and previous assessments (2016 and 2011) for countries that participated in both. For each country, trend item statistics included the percentage of students in each score category (or response option for multiple-choice items) for each assessment, as well as the difficulty of the item and the percent correct by gender. In reviewing these item statistics, the aim was to detect any unusual changes in item difficulties between administrations, which might indicate a problem in using the item to measure trends.

Exhibit 10.4: Example Item Statistics for a PIRLS 2016 Trend Item

Progress in International Reading Literacy Study - PIRLS 2016 Assessment Results
Trend Achievement Data Almanac for Literary Experience Items (Weighted)

The Empty Pot: Literary Experience / Evaluate and Critique Content and Textual Elements
R31M15M: What Emperor values in a person - 1 Point - Key: B

COUNTRY	YEAR	N	DIFF %	A %	B %	C %	D %	OMITTED %	NOT REACHED %	1.GIRL % RIGHT	2.BOY % RIGHT
Australia	2011	1206	54.9	11.7	54.9	5.7	26.6	0.3	0.9	59.7	50.7
	2016	1056	55.8	9.1	55.8	3.6	28.7	2.0	0.8	59.5	52.0
Austria	2011	937	64.7	8.7	64.7	2.7	22.3	1.3	0.3	66.9	62.8
	2016	726	63.9	8.2	63.9	1.2	23.5	2.6	0.6	68.0	59.9
Azerbaijan	2011	954	53.0	8.3	53.0	7.2	21.7	3.3	6.6	53.8	52.3
	2016	995	51.9	11.2	51.9	5.5	27.4	1.9	2.0	54.4	49.7
Belgium (French)	2011	731	43.6	13.2	43.6	3.0	36.9	1.7	1.6	43.1	44.1
	2016	774	48.1	10.4	48.1	3.0	34.3	2.4	1.8	50.8	45.4
Bulgaria	2011	1053	72.4	5.3	72.4	3.7	15.5	0.5	2.7	75.5	69.5
	2016	724	79.8	3.9	79.8	3.8	11.8	0.3	0.3	82.5	77.1
Canada	2011	4592	66.2	8.6	66.2	3.4	20.2	0.6	0.9	65.4	67.1
	2016	3000	63.0	7.7	63.0	3.6	21.8	1.0	3.1	68.7	57.6
Chinese Taipei	2011	857	82.5	2.9	82.5	1.8	12.1	0.7	0.0	82.9	82.3
	2016	718	89.0	1.6	89.0	1.0	8.1	0.3	0.1	89.5	88.4
Czech Republic	2011	907	57.8	1.8	57.8	4.0	34.5	1.2	0.7	56.6	58.9
	2016	927	63.7	1.5	63.7	4.2	27.3	2.0	1.4	62.1	65.1
Denmark	2011	902	71.9	2.9	71.9	2.7	20.2	0.8	1.6	74.0	69.7
	2016	583	77.7	1.8	77.7	1.7	15.4	0.5	2.9	78.3	77.0
England	2011	780	62.4	9.5	62.4	3.5	23.6	0.6	0.5	62.6	62.2
	2016	834	70.5	7.5	70.5	2.5	18.4	0.8	0.4	73.9	67.1
Finland	2011	920	76.1	2.5	76.1	1.7	19.0	0.3	0.4	78.5	73.8
	2016	809	72.4	2.9	72.4	1.6	21.7	0.5	0.8	78.0	66.2
France	2011	876	52.1	10.2	52.1	2.8	31.0	2.6	1.3	48.6	55.2
	2016	789	54.0	9.0	54.0	3.6	28.2	2.9	2.2	56.6	51.1
Georgia	2011	951	52.4	7.2	52.4	5.5	30.6	0.8	3.4	54.1	50.8
	2016	956	52.3	7.9	52.3	6.0	28.8	1.2	3.9	58.7	46.1
Germany	2011	798	62.8	8.2	62.8	1.8	24.4	1.8	1.0	62.4	63.3
	2016	649	63.8	6.0	63.8	2.5	23.6	3.2	0.9	64.3	63.2
Hong Kong SAR	2011	770	95.3	1.4	95.3	0.8	2.3	0.2	0.0	95.4	95.2
	2016	557	94.5	2.0	94.5	0.5	2.3	0.2	0.5	96.7	92.5
Hungary	2011	1026	72.3	6.1	72.3	2.9	16.0	1.5	1.3	74.9	69.7
	2016	766	77.1	4.1	77.1	1.6	14.1	1.1	2.0	78.5	75.7
Iran, Islamic Rep. of	2011	1148	28.4	16.0	28.4	13.0	37.4	1.5	3.7	29.4	27.4
	2016	724	27.0	15.4	27.0	15.4	33.2	3.1	5.9	30.4	23.9

DIFF = Percent correct
Because of missing gender information, some totals may appear inconsistent.

Exhibit 10.4: Example Item Statistics for a PIRLS 2016 Trend Item (Continued)

Progress in International Reading Literacy Study - PIRLS 2016 Assessment Results
Trend Achievement Data Almanac for Literary Experience Items (Weighted)

The Empty Pot: Literary Experience / Evaluate and Critique Content and Textual Elements
R31M15M: What Emperor values in a person - 1 Point - Key: B

COUNTRY	YEAR	N	DIFF %	A %	B %	C %	D %	OMITTED %	NOT REACHED %	1.GIRL % RIGHT	2.BOY % RIGHT
Ireland	2011	903	60.4	9.2	60.4	3.1	26.0	0.6	0.7	68.0	52.8
	2016	762	67.4	7.9	67.4	1.2	22.6	0.5	0.4	71.3	63.6
Israel	2011	835	63.1	9.4	63.1	5.2	18.6	0.8	2.9	63.3	62.9
	2016	670	69.1	8.9	69.1	5.2	12.2	1.5	3.1	73.2	65.4
Italy	2011	830	69.9	5.1	69.9	4.1	18.3	1.8	0.8	72.2	67.8
	2016	649	75.2	4.4	75.2	2.5	15.4	1.6	0.9	77.4	73.1
Lithuania	2011	913	70.6	5.4	70.6	2.3	21.1	0.3	0.3	74.8	66.9
	2016	728	71.9	6.8	71.9	2.3	18.3	0.3	0.4	75.0	68.5
Malta	2011	717	41.1	17.0	41.1	8.9	30.3	0.8	2.0	48.0	35.0
	2016	605	50.0	11.7	50.0	8.8	25.2	1.6	2.8	49.1	50.7
Morocco	2011	1460	28.4	15.8	28.4	17.5	24.0	4.3	9.9	25.2	31.6
	2016	895	28.2	15.4	28.2	16.1	30.1	2.2	8.1	29.8	26.7
Netherlands	2011	803	79.8	5.3	79.8	1.5	13.0	0.3	0.3	79.7	79.9
	2016	700	80.0	6.2	80.0	1.6	10.8	0.8	0.7	79.8	80.3
New Zealand	2011	1136	57.2	10.9	57.2	5.0	24.6	1.0	1.2	60.9	53.5
	2016	934	57.8	10.5	57.8	4.8	24.0	1.0	1.9	63.0	52.8
Northern Ireland	2011	707	65.2	9.9	65.2	2.4	21.6	0.6	0.2	69.9	60.5
	2016	610	70.6	9.1	70.6	1.5	18.0	0.7	0.1	73.1	67.8
Norway (4)	2011	633	59.2	5.1	59.2	3.4	27.1	2.6	2.5	62.1	55.8
	2016	723	63.3	2.7	63.3	2.7	23.0	1.9	6.3	66.3	60.8
Oman	2011	2041	25.3	20.4	25.3	16.0	29.2	3.4	5.8	24.8	25.6
	2016	1539	32.8	18.9	32.8	13.3	29.7	1.8	3.5	32.4	33.3
Portugal	2011	815	61.4	5.3	61.4	3.0	28.6	0.9	0.9	61.0	61.7
	2016	758	62.2	4.3	62.2	2.3	28.5	0.8	2.0	62.5	61.8
Qatar	2011	805	29.5	20.8	29.5	13.9	31.0	2.1	2.7	27.4	31.5
	2016	1502	44.0	16.1	44.0	9.5	25.5	1.0	3.9	44.9	43.1
Russian Federation	2011	888	74.9	3.3	74.9	1.7	19.3	0.7	0.1	76.2	73.6
	2016	757	80.1	4.1	80.1	1.8	12.9	0.5	0.5	79.5	80.8
Saudi Arabia	2011	898	34.2	17.7	34.2	16.2	28.2	1.5	2.2	30.7	38.0
	2016	783	42.9	17.7	42.9	14.6	18.4	1.2	5.2	43.0	42.8
Singapore	2011	1254	84.5	6.5	84.5	1.8	6.8	0.4	0.1	86.8	82.3
	2016	1083	85.4	6.4	85.4	0.9	6.9	0.3	0.1	86.8	84.1
Slovak Republic	2011	1119	68.2	4.3	68.2	5.1	20.7	1.2	0.5	68.2	68.1
	2016	907	61.7	6.8	61.7	6.0	23.8	0.8	0.9	64.0	59.3

DIFF = Percent correct
Because of missing gender information, some totals may appear inconsistent.

Exhibit 10.4: Example Item Statistics for a PIRLS 2016 Trend Item (Continued)

Progress in International Reading Literacy Study - PIRLS 2016 Assessment Results
Trend Achievement Data Almanac for Literary Experience Items (Weighted)

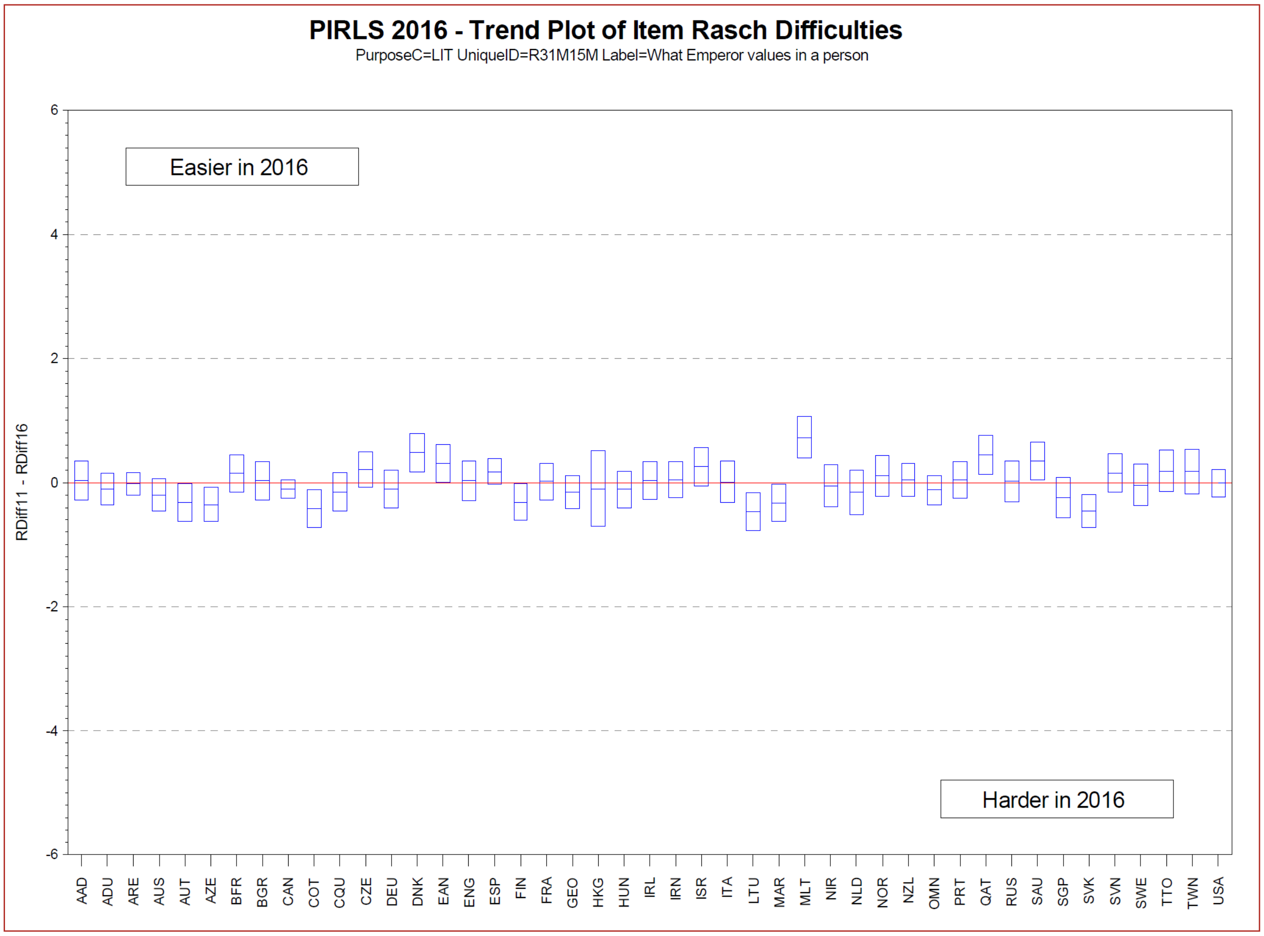
The Empty Pot: Literary Experience / Evaluate and Critique Content and Textual Elements
R31M15M: What Emperor values in a person - 1 Point - Key: B

COUNTRY	YEAR	N	DIFF %	A %	B %	C %	D %	OMITTED %	NOT REACHED %	1.GIRL % RIGHT	2.BOY % RIGHT
Slovenia	2011	881	67.0	3.2	67.0	1.6	26.4	0.7	1.1	73.4	61.1
	2016	751	75.9	5.0	75.9	1.8	15.9	0.4	1.0	77.5	74.5
Spain	2011	1692	53.2	8.0	53.2	3.9	31.5	0.7	2.7	53.9	52.4
	2016	2431	65.9	7.3	65.9	2.2	21.6	0.9	2.1	67.6	64.1
Sweden	2011	905	75.7	1.8	75.7	3.1	18.7	0.5	0.2	76.5	74.7
	2016	749	81.4	2.9	81.4	1.5	11.7	0.8	1.6	81.7	81.1
Trinidad and Tobago	2011	787	39.9	13.9	39.9	5.4	35.5	2.8	2.5	43.4	36.7
	2016	687	47.6	13.7	47.6	6.1	26.4	2.0	4.3	47.9	47.2
United Arab Emirates	2011	2894	39.0	17.8	39.0	11.2	27.7	1.7	2.6	38.4	39.6
	2016	2730	44.1	17.3	44.1	8.2	24.7	3.1	2.6	45.9	42.3
United States	2011	2532	67.0	10.7	67.0	2.4	17.9	0.4	1.6	68.2	65.7
	2016	748	67.2	8.9	67.2	2.6	19.1	0.4	1.8	71.5	63.5
International Avg. (40)	2011	45856	59.6	8.8	59.6	5.1	23.5	1.2	1.8	60.9	58.3
	2016	38288	63.2	8.1	63.2	4.5	20.8	1.3	2.1	65.3	61.1
Ontario, Canada	2011	903	64.1	9.1	64.1	2.9	22.4	0.4	1.1	62.0	66.4
	2016	693	58.7	9.1	58.7	4.7	24.2	1.0	2.3	63.6	54.8
Quebec, Canada	2011	848	72.3	8.3	72.3	2.0	16.0	0.5	0.8	72.1	72.5
	2016	527	74.3	6.8	74.3	1.8	14.3	1.0	1.8	81.0	66.1
Andalusia, Spain	2011	855	57.1	7.1	57.1	2.8	29.7	1.5	1.8	60.6	53.5
	2016	691	67.3	5.8	67.3	2.2	22.1	0.8	1.7	65.7	68.8
Abu Dhabi, UAE	2011	817	35.9	19.6	35.9	14.6	26.1	1.4	2.3	33.9	38.0
	2016	690	37.8	17.3	37.8	9.6	26.0	5.6	3.8	41.0	34.5
Dubai, UAE	2011	1199	47.2	13.0	47.2	5.7	29.9	1.9	2.3	47.8	46.7
	2016	1306	59.1	12.3	59.1	4.2	21.5	1.4	1.5	59.4	58.8

DIFF = Percent correct
Because of missing gender information, some totals may appear inconsistent.

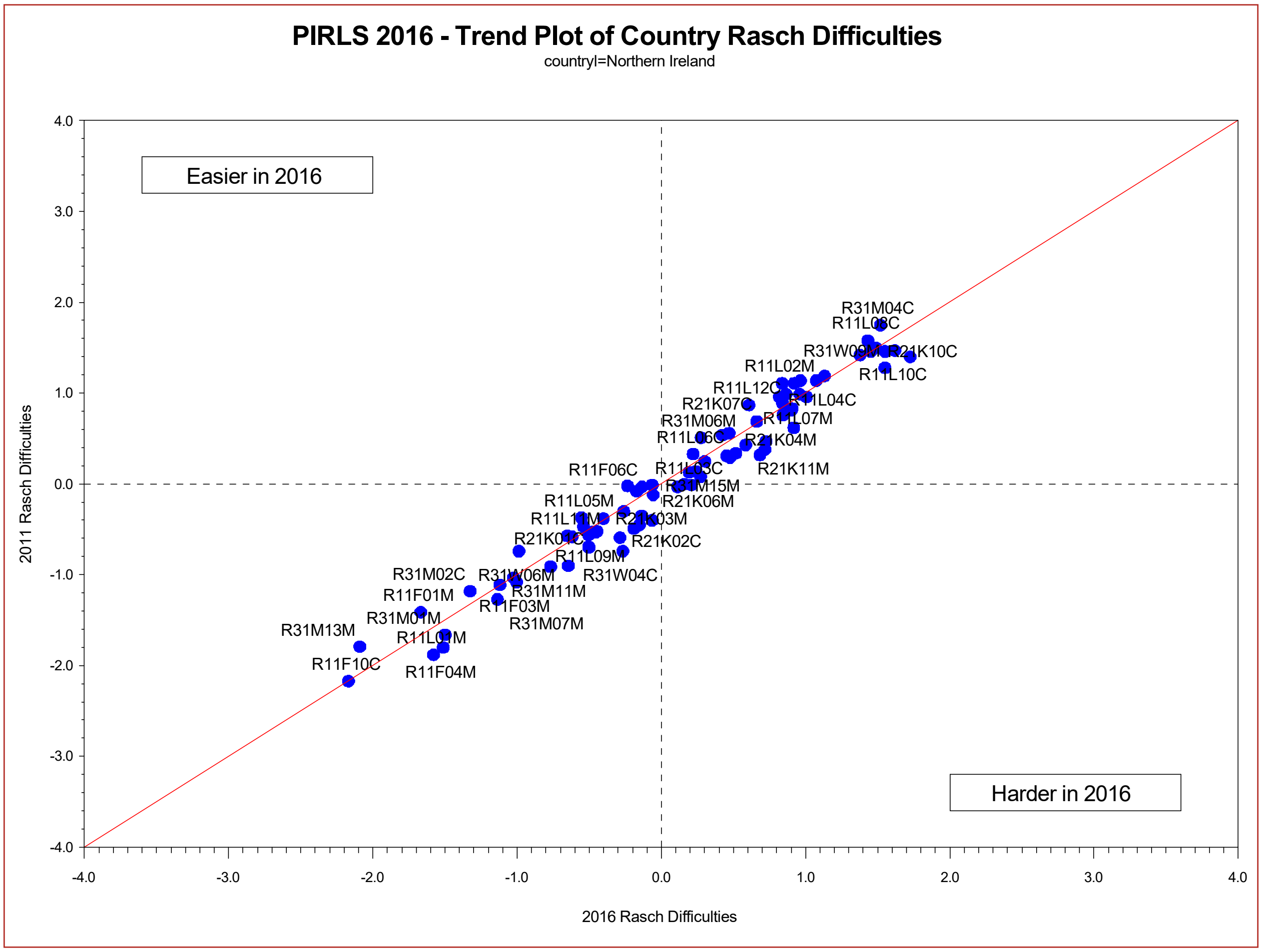
Although some changes in item difficulties were anticipated as countries' overall achievement may have improved or declined, items were noted if the difference between the Rasch difficulties across the two assessments for a particular country was greater than 2 logits. The TIMSS & PIRLS International Study Center used two different graphical displays to examine the differences in item difficulties. The first of these, shown for an example item in Exhibit 10.5, displays the difference in Rasch item difficulty of the item between 2016 and 2011 for each country. A positive difference for a country indicates that the item was relatively easier in 2016, and a negative difference indicates that the item was relatively more difficult.

Exhibit 10.5: Example Plot of Differences in Rasch Item Difficulties Between 2016 and 2011 for a PIRLS 2016 Trend Item



The second graphical display, presented in Exhibit 10.6, shows the performance of a given country on all trend items simultaneously. For each country, the graph plots the 2016 Rasch difficulty of every trend item against its Rasch difficulty in 2011. Where there were no differences between the difficulties in the two successive administrations, the data points aligned on or near the diagonal.

Exhibit 10.6: Example Plot of Rasch Item Difficulties Across PIRLS 2016 Trend Items by Country



Reliability

Documenting the reliability of the PIRLS 2016 assessments was a critical quality control step in reviewing the items. As one indicator of reliability, the review considered Cronbach's Alpha coefficient of reliability calculated at the assessment booklet level. Secondly, the scoring of the constructed response items had to meet specific reliability criteria in terms of consistent within-country scoring, cross-country scoring, and across assessment or trend scoring.

Test Reliability

Exhibit 10.7 displays the PIRLS, PIRLS Literacy, and ePIRLS test reliability coefficients for every country, respectively. These coefficients are the median Cronbach's alpha reliability across all PIRLS 2016 assessment booklets. In general, reliabilities were relatively high. For PIRLS, the international median reliability (the median of the reliability coefficients for all countries) was 0.83. The international median reliability for PIRLS Numeracy was 0.92, whereas the international median reliability for ePIRLS was 0.92.

Exhibit 10.7: Cronbach's Alpha Reliability Coefficient – PIRLS 2016

Country	Reliability Coefficient		
	PIRLS	PIRLS Literacy	ePIRLS
Australia	0.91	—	—
Austria	0.86	—	—
Azerbaijan	0.89	—	—
Bahrain	0.91	—	—
Belgium (Flemish)	0.86	—	—
Belgium (French)	0.87	—	—
Bulgaria	0.91	—	—
Canada	0.89	—	0.90
Chile	0.90	—	—
Chinese Taipei	0.87	—	0.90
Czech Republic	0.88	—	—
Denmark	0.88	—	0.90
Egypt	—	0.92	—
England	0.90	—	—
Finland	0.88	—	—
France	0.88	—	—
Georgia	0.89	—	0.90
Germany	0.90	—	—
Hong Kong SAR	0.85	—	—
Hungary	0.89	—	—
Iran, Islamic Rep. of	0.90	0.92	—
Ireland	0.89	—	0.90
Israel	0.92	—	0.92
Italy	0.87	—	0.89
Kazakhstan	0.86	—	—
Kuwait	—	0.90	—
Latvia	0.86	—	—
Lithuania	0.88	—	—
Macao SAR	0.87	—	—
Malta	0.89	—	—

Exhibit 10.7: Cronbach's Alpha Reliability Coefficient – PIRLS 2016 (Continued)

Country	Reliability Coefficient		
	PIRLS	PIRLS Literacy	ePIRLS
Morocco	0.86	0.91	—
Netherlands	0.86	—	—
New Zealand	0.92	—	—
Northern Ireland	0.90	—	—
Norway	0.87	—	0.89
Oman	0.91	—	—
Poland	0.88	—	—
Portugal	0.87	—	0.89
Qatar	0.92	—	—
Russian Federation	0.87	—	—
Saudi Arabia	0.90	—	—
Singapore	0.91	—	0.92
Slovak Republic	0.90	—	—
Slovenia	0.89	—	0.90
South Africa	—	0.90	—
Spain	0.87	—	—
Sweden	0.88	—	0.90
Trinidad and Tobago	0.92	—	—
United Arab Emirates	0.93	—	0.93
United States	0.90	—	0.91
International Median	**0.89**	**0.91**	**0.90**
Benchmarking Participants			
Buenos Aires, Argentina	0.90	—	—
Ontario, Canada	0.90	—	—
Quebec, Canada	0.85	—	—
Denmark (3)	—	0.88	—
Norway (4)	0.88	—	—
Moscow City, Russian Fed.	0.83	—	—
Eng/Afr/Zulu - RSA (5)	0.91	—	—
Andalusia, Spain	0.86	—	—
Madrid, Spain	0.85	—	—
Abu Dhabi, UAE	0.92	—	0.93
Dubai, UAE	0.92	—	0.93

Scoring Reliability for Constructed Response Items

A sizeable proportion of the items in the PIRLS 2016 assessments were constructed response items, comprising about half of the assessment score points. An essential requirement for use of such items is that they be reliably scored by all participants. That is, a particular student response should receive the same score, regardless of the scorer. In conducting PIRLS 2016, measures taken to ensure that the constructed response items were scored reliably in all countries included developing scoring guides for each constructed response question (that provided descriptions of acceptable responses for each score point value) and providing extensive training in the application of the scoring guides. See Chapter 1: Developing the PIRLS 2016 Achievement Items for more information on the scoring guides and see Chapter 6: Survey Operations for PIRLS 2016 for information on the scoring process.

Within-Country Scoring Reliability

To gather and document information about the within-country agreement among scorers for PIRLS 2016, a random sample of approximately 25 percent of the assessment booklets was selected to be scored independently by two scorers. The inter-scorer agreement for each item in each country was examined as part of the item review process. Exact percent agreement across items was high on average across countries—96 percent or above, on average internationally. See Appendix 10A for the average and range of the within-country percentage of correctness score agreement across all items. The PIRLS Within-Country Scoring Reliability documents also provide the average and range of the within-country percentage of diagnostic score agreement.

Trend Item Scoring Reliability

The TIMSS & PIRLS International Study Center also took steps to show that the 2016 constructed response items used in PIRLS 2011 were scored in the same way in both assessments. In anticipation of this, countries that participated in PIRLS 2011 sent samples of scored student booklets from the 2011 data collections to IEA Hamburg, where they were digitally scanned and stored for later use. As a check on scoring consistency from one administration to the next, staff members working in each country on scoring the 2016 data were asked also to score these 2011 responses using the Trend Reliability Scoring Software developed by IEA Hamburg. Each country scored 200 responses for 22 PIRLS reading items (South Africa scored 24 PIRLS Literacy reading items for their fourth grade sample).

There was a very high degree of scoring consistency in PIRLS 2016. The exact agreement between the scores awarded in 2011 and those given by the 2016 scorers was 95 percent on average internationally. The average and range of scoring consistency over time can be found in Appendix 10B.

Cross-Country Scoring Reliability Study

It also was important to document the consistency of scoring across countries. Because of the many different languages in use in PIRLS 2016, establishing the reliability of constructed response scoring across all countries was not feasible. However, the TIMSS & PIRLS International Study Center did conduct a cross-country study of scoring reliability among Northern Hemisphere countries that had scorers who were proficient in English. A sample of student responses was provided by the English-speaking Southern Hemisphere countries. Cross-country scoring included 200 student responses for 22 PIRLS reading items. This set of student responses in English was then scored independently in each country that had two scorers proficient in English, using the Cross-Country Scoring Reliability Software provided by IEA Hamburg. In all, scorers from 44 countries and four benchmarking participants took part in the process. Making all possible comparisons among scorers gave a total of 1,128 possible comparisons for each student response to each item, and resulted in more than 225,600 total comparisons when aggregated across all 200 student responses to any given item.

Agreement across countries was defined in terms of the percentage of these comparisons that were in exact agreement. On average internationally, scorer reliability across countries in PIRLS 2016 was high, with an exact agreement in the scores awarded of 85 percent on average internationally. See Appendix 10C for the results of the cross-country scoring reliability study.

Item Review Procedures

Using the information from the comprehensive collection of item analyses and reliability data that were computed and summarized for PIRLS 2016, the TIMSS & PIRLS International Study Center thoroughly reviewed all item statistics for every participating country and benchmarking participant to ensure that the items were performing comparably across countries. In particular, items with the following problems were considered for possible deletion from the international database:

- An error was detected during translation verification but was not corrected before test administration
- Data checking revealed a multiple-choice item with more or fewer options than in the international version
- The item analysis showed the item to have a negative biserial, or, for an item with more than 1 score point, point biserials that did not increase with each score level
- The item-by-country interaction results showed a very large negative interaction for a particular country

- For constructed response items, the within-country scoring reliability data showed an agreement of less than 70 percent
- For trend items, an item performed substantially differently in 2016 compared to the PIRLS 2011 administration, or an item was not included in the previous assessment for a particular country

When item statistics indicated a problem with an item, translation verification documentation was used as an aid in checking the test booklets. If a question remained about potential translation or cultural issues, however, then the National Research Coordinator was consulted before deciding how the item should be treated.

The checking of the PIRLS 2016 achievement data involved review of almost 400 items and resulted in the detection of very few items that were inappropriate for international comparisons. The items found to be problematic during the review process primarily had issues related to translation or printing problems. See Appendix 10D: Country Adaptations to Items and Item Scoring for a list of deleted items, as well as a list of recodes made to constructed response item codes. There also were a number of items in each study that were combined, or derived, for scoring purposes. See Appendix 10E for details about how score points were awarded for each derived item.

Appendix 10A: PIRLS 2016 Within-Country Scoring Reliability for the Constructed Response Items

PIRLS 2016 Within-Country Scoring Reliability for the Constructed Response Items

Country	PIRLS		
	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
Australia	92	68	100
Austria	96	88	100
Azerbaijan	98	93	100
Bahrain	96	89	100
Belgium (Flemish)	95	84	100
Belgium (French)	99	95	100
Bulgaria	96	84	100
Canada	89	68	100
Chile	98	94	100
Chinese Taipei	98	91	100
Czech Republic	100	98	100
Denmark	90	62	100
England	95	75	100
Finland	96	79	100
France	94	83	100
Georgia	91	72	100
Germany	93	72	100
Hong Kong SAR	97	84	100
Hungary	97	91	100
Iran, Islamic Rep. of	95	85	100
Ireland	99	94	100
Israel	96	87	100
Italy	95	86	100
Kazakhstan	99	96	100
Latvia	96	83	100
Lithuania	99	97	100
Macao SAR	99	96	100
Malta	91	76	100
Morocco	89	68	99
Netherlands	96	82	100

PIRLS 2016 Within-Country Scoring Reliability for the Constructed Response Items (Continued)

Country	PIRLS		
	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
New Zealand	95	79	100
Northern Ireland	100	100	100
Norway	97	91	100
Oman	94	85	100
Poland	94	80	100
Portugal	98	90	100
Qatar	98	93	100
Russian Federation	99	94	100
Saudi Arabia	98	92	100
Singapore	100	99	100
Slovak Republic	98	91	100
Slovenia	97	82	100
Spain	98	90	100
Sweden	96	88	100
Trinidad and Tobago	89	66	100
United Arab Emirates	93	86	100
United States	97	91	100
International Avg.	**96**	**86**	**100**
Benchmarking Participants			
Buenos Aires, Argentina	93	79	100
Ontario, Canada	88	70	100
Quebec, Canada	89	59	100
Moscow City, Russian Fed.	98	90	100
Eng/Afr/Zulu - RSA (5)	92	76	100
Andalusia, Spain	98	91	100
Madrid, Spain	98	87	100
Abu Dhabi, UAE	93	84	100
Dubai, UAE	93	86	100

PIRLS Literacy 2016 Within-Country Scoring Reliability for the PIRLS Literacy Constructed Response Items

Country	PIRLS Literacy		
	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
Egypt	97	88	100
Iran, Islamic Rep. of	96	76	100
Kuwait	90	61	100
Morocco	89	33	100
South Africa	94	83	100
International Avg.	**93**	**68**	**100**
Benchmarking Participant			
Denmark (3)	95	68	100

ePIRLS 2016 Within-Country Scoring Reliability for the ePIRLS Constructed Response Items

Country	ePIRLS		
	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
Canada	92	79	99
Chinese Taipei	96	90	100
Denmark	91	72	99
Georgia	94	84	100
Ireland	95	90	100
Israel	95	90	100
Italy	95	88	100
Norway	98	96	100
Portugal	95	87	100
Singapore	100	100	100
Slovenia	90	70	100
Sweden	95	86	100
United Arab Emirates	94	87	100
United States	94	86	100
International Avg.	**95**	**86**	**100**
Benchmarking Participants			
Abu Dhabi, UAE	94	86	100
Dubai, UAE	93	85	100

Appendix 10B: PIRLS 2016 Trend Scoring Reliability for the Constructed Response Items

PIRLS 2016 Trend Scoring Reliability for the Constructed Response Items

Country	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
Australia	95	80	100
Austria	96	82	100
Azerbaijan	92	66	100
Belgium (French)	97	87	100
Bulgaria	96	81	100
Canada	94	79	100
Chinese Taipei	95	81	100
Czech Republic	96	80	100
Denmark	95	78	100
England	96	80	100
Finland	95	78	100
France	93	69	100
Georgia	93	76	100
Germany	96	85	100
Hong Kong SAR	98	88	100
Hungary	95	78	100
Iran, Islamic Rep. of	95	82	100
Ireland	96	86	100
Israel	95	78	100
Italy	94	82	100
Lithuania	97	90	100
Netherlands	94	69	100
New Zealand	96	80	100
Northern Ireland	96	83	100
Norway	96	87	100
Oman	95	82	100
Poland	96	82	100
Portugal	92	66	100
Qatar	91	62	100
Russian Federation	96	83	100
Singapore	96	83	100
Slovak Republic	94	82	100

PIRLS 2016 Trend Scoring Reliability for the Constructed Response Items (Continued)

Country	Average of Percent Agreement Across Items	Range of Percent Agreement	
		Minimum	Maximum
South Africa	93	72	100
Spain	92	72	100
Sweden	95	78	100
Trinidad and Tobago	92	73	100
United Arab Emirates	93	56	100
United States	94	74	100
International Avg.	**95**	**78**	**100**
Benchmarking Participants			
Eng/Afr/Zulu - RSA (5)	91	65	100
Dubai, UAE	90	51	100

Appendix 10C: PIRLS 2016 Cross-Country Scoring Reliability for the Constructed Response Items

PIRLS 2016 Cross-Country Scoring Reliability for the Constructed Response Items

Item Label	Total Valid Comparisons	Percent Exact Agreement
Empty Pot R31M02C	214,879	97
Empty Pot R31M04C	204,588	88
Empty Pot R31M09C	212,582	86
Empty Pot R31M10C	216,460	92
Empty Pot R31M16C	216,989	92
Honey R31W01C	221,321	94
Honey R31W02C	211,896	78
Honey R31W04C	213,069	96
Honey R31W11C	217,978	97
Honey R31W13C	217,192	84
Sharks R21K01C	214,490	81
Sharks R21K02C	216,596	93
Sharks R21K05C	212,590	87
Sharks R21K07C	208,487	81
Sharks R21K10C	213,352	82
Sharks R21K12C	214,311	77
Shiny Straw R21Y03C	210,586	89
Shiny Straw R21Y09C	215,727	82
Shiny Straw R21Y10C	212,668	78
Shiny Straw R21Y12C	214,658	80
Shiny Straw R21Y13C	215,811	65
Shiny Straw R21Y14C	209,761	73
Average Percent Agreement		**85**

Appendix 10D: Country Adaptations to PIRLS 2016 Items and Item Scoring

Country Adaptations to PIRLS 2016 Items and Item Scoring

PIRLS and PIRLS Literacy
Deleted Items
MALTA
The Green Sea Turtle's Journey of a Lifetime Item 15, R41T15M (Negative discrimination)
How Did We Learn to Fly? Item 2, L21E02C (Translation error)
NETHERLANDS
Sharks Item 4, R21K04M (Negative discrimination)
NORWAY
Sharks Item 2, R21K02C (Printing error)
Shiny Straw Item 3, R21Y03C (Printing error)
Empty Pot Item 7, R31M07M (Translation error)
SAUDI ARABIA
Oliver and the Griffin Item 6, R41O06M (Low discrimination)
SINGAPORE
The Green Sea Turtle's Journey of a Lifetime Item 11, R41T11C (Scoring error)
SLOVAK REPUBLIC
Empty Pot Item 4, R31M04C (Translation error)
Constructed Response Items with Category Recoding
Icelandic Horses Item 15, R41I15C (Recoded from 2 into 1)
African Rhinos and Oxpecker Birds Item 17, L21C17C (Recoded from 2 into 1)
Flowers on the Roof Item 12, R11F12C (Recoded from 3 into 2)
ePIRLS
Deleted Items
GEORGIA
Rainforests Item 6, E11R06C (Missing data)
Constructed Response Items with Category Recoding
Zebra and Wildebeest Migration Item 12, E11Z12C (Recoded 2 to 1)

Appendix 10E: Derived Items in PIRLS 2016

Derived Items in PIRLS 2016

PIRLS and PIRLS Literacy
Where's the Honey? Item 7, R31W07C – Item parts A, B, and C are combined to create a 3-point item, where 3 score points are awarded if all parts are correct, 2 score points are awarded if two parts are correct, and 1 score point is awarded if only one part is correct
Empty Pot Item 17, R31M17C – Item parts A, B, and C are combined to create a 3-point item, where 3 score points are awarded if all parts are correct, 2 score points are awarded if two parts are correct, and 1 score point is awarded if only one part is correct
Ants Item 12, L11A12CZ – Item parts A, B, and C are summed to create a 3-point item
Ants Item 13, L11A13CZ – Item parts B–E are combined to create a 2-point item, where 2 points are awarded if all 4 parts are correct, 1 point is awarded if 3 parts are correct, and 0 points are awarded if 2 or fewer parts are correct
ePIRLS
Mars Item 16, E11M16C – Item parts A through D are combined to create a 2-point item, where 2 points are awarded if all 4 parts are correct, 1 point is awarded if 3 parts are correct, and 0 points are awarded if 2 or fewer parts are correct
Rainforests Item 3, E11R03C – Item parts A through D are combined to create a 2-point item, where 2 points are awarded if all 4 parts are correct, 1 point is awarded if 3 parts are correct, and 0 points are awarded if 2 or fewer parts are correct
Rainforests Item 7, E11R07C – Item parts A through D are combined to create a 2-point item, where 2 points are awarded if all 4 parts are correct, 1 point is awarded if 3 parts are correct, and 0 points are awarded if 2 or fewer parts are correct
Zebra and Wildebeest Migration item 20, E11Z20C – Item parts A through D are combined to create a 2-point item, where 2 points are awarded if all 4 parts are correct, 1 point is awarded if 3 parts are correct, and 0 points are awarded if 2 or fewer parts are correct
The Legend of Troy Item 18, E11T18C – Item parts A, B, and D are combined to create a 1-point item, where 1 point is awarded if all 3 parts are correct and 0 points are awarded if 2 or fewer parts are correct

CHAPTER 11

PIRLS 2016 Achievement Scaling Methodology[1]

The PIRLS approach to scaling the achievement data, based on item response theory (IRT) scaling with marginal estimation, was developed originally by Educational Testing Service for use in the U.S. National Assessment of Educational Progress (NAEP). It is based on psychometric models that were first used in the field of educational measurement in the 1950s and have become popular since the 1970s for use in large-scale surveys, test construction, and computer adaptive testing.[2]

Three distinct IRT models, depending on item type and scoring procedure, were used in the analysis of the PIRLS 2016 assessment data. Each is a "latent variable" model that describes the probability that a student will respond in a specific way to an item in terms of the student's proficiency, which is an unobserved or "latent" trait, and various characteristics (or "parameters") of the item. A three-parameter model was used with multiple-choice items, which were scored as correct or incorrect, and a two-parameter model for constructed response items with just two response options, which also were scored as correct or incorrect. Since each of these item types has just two response categories, they are known as dichotomous items. A partial credit model was used with polytomous constructed response items, i.e., those with more than two response options.

Two- and Three-Parameter IRT Models for Dichotomous Items

The fundamental equation of the three-parameter logistic (3PL) model gives the probability that a student whose proficiency on a scale k is characterized by the unobservable variable θ_k will respond correctly to item i as:

$$P\left(x_i=1 \middle| \theta_k, a_i, b_i, c_i\right) = c_i + \frac{1-c_i}{1+\exp\left(-1.7 \cdot a_i \cdot (\theta_k - b_i)\right)} \equiv P_{i,1}\left(\theta_k\right) \qquad (11.1)$$

1 This description of the PIRLS achievement scaling methodology has been adapted with permission from the TIMSS 1999 Technical Report (Yamamoto and Kulick, 2000).

2 For a description of IRT scaling see Birnbaum (1968); Lord and Novick (1968); Lord (1980); Van Der Linden and Hambleton (1996). The theoretical underpinning of the multiple imputation methodology was developed by Rubin (1987), applied to large-scale assessment by Mislevy (1991), and studied further by Mislevy, Johnson, and Muraki (1992) and Beaton and Johnson (1992). For a recent overview, see von Davier and Sinharay (2014) and von Davier (2014). The procedures used in PIRLS have been used in several other large-scale surveys, including the U.S. National Assessment of Educational Progress (NAEP), the U.S. National Adult Literacy Survey (NALS), the International Adult Literacy Survey (IALS), and the International Adult Literacy and Life Skills Survey (IALLS).

where

x_i is the response to item i, 1 if correct and 0 if incorrect;

θ_k is the proficiency of a student on a scale k (note that a student with higher proficiency has a greater probability of responding correctly);

a_i is the slope parameter of item i, characterizing its discriminating power;

b_i is the location parameter of item i, characterizing its difficulty;

c_i is the lower asymptote parameter of item i, reflecting the chances of students with very low proficiency selecting the correct answer.

The probability of an incorrect response to the item is defined as:

$$P_{i,0} = P\left(x_i = 0 \middle| \theta_k, a_i, b_i, c_i\right) = 1 - P_{i,1}\left(\theta_k\right) \qquad (11.2)$$

The two-parameter logistic (2PL) model was used for the constructed response items that were scored as either correct or incorrect. The form of the 2PL model is the same as Equations (11.1) and (11.2) with the c_i parameter fixed at zero.

IRT Model for Polytomous Items

In PIRLS, constructed response items requiring an extended response were scored for partial credit, with 0, 1, 2, and 3 as the possible score levels. These polytomous items were scaled using a generalized partial credit model (Muraki, 1992). The fundamental equation of this model gives the probability that a student with proficiency θ_k on scale k will have, for the i^{th} item, a response x_i that is scored in the l^{th} of m_i ordered score categories as:

$$P\left(x_i = l \middle| \theta_k, a_i, b_i, d_{i,1}, \cdots, d_{i,m_i-1}\right) = \frac{\exp\left(\sum_{v=0}^{l} 1.7 \cdot a_i \cdot \left(\theta_k - b_i + d_{i,v}\right)\right)}{\sum_{g=0}^{m_i-1} \exp\left(\sum_{v=0}^{g} 1.7 \cdot a_i \cdot \left(\theta_k - b_i + d_{i,v}\right)\right)} = P_{i,l}\left(\theta_k\right) \qquad (11.3)$$

where

m_i is the number of response categories for item i;

x_i is the response to item i, ranging between 0 and $m_i - 1$;

θ_k is the proficiency of a student on a scale k;

a_i is the slope parameter of item i;

b_i is its location parameter, characterizing its difficulty;

$d_{i,l}$ is the category l threshold parameter.

The indeterminacy of model parameters in the polytomous model is resolved by setting $d_{i,0} = 0$ and $\sum_{j=1}^{m_i-1} d_{i,j} = 0$.

For all of the IRT models there is a linear indeterminacy between the values of item parameters and proficiency parameters, i.e., mathematically equivalent but different values of item parameters can be estimated on an arbitrarily linearly transformed proficiency scale. This linear indeterminacy can be resolved by setting the origin and unit size of the proficiency scale to arbitrary constants, such as a mean of 500 and a standard deviation of 100, as was done originally for PIRLS 2001. The indeterminacy is most apparent when the scale is set for the first time.

IRT modeling relies on a number of assumptions, the most important being conditional independence. Under this assumption, item response probabilities depend only on θ_k (a measure of a student's proficiency) and the specified parameters of the item, and are unaffected by the demographic characteristics or unique experiences of the students, the data collection conditions, or the other items presented in the test. Under this assumption, the joint probability of a particular response pattern x across a set of n items is given by:

$$P\left(x \middle| \theta_k, \textit{item parameters}\right) = \prod_{i=1}^{n} \prod_{l=0}^{m_i - 1} P_{i,l}\left(\theta_k\right)^{u_{i,l}} \tag{11.4}$$

where $P_{i,l}(\theta_k)$ is of the form appropriate to the type of item (dichotomous or polytomous), m_i is equal to 2 for dichotomously scored items, and $u_{i,l}$ is an indicator variable defined as:

$$u_{i,l} = \begin{cases} 1 \textit{ if response is } x_i \textit{ is in category } l; \\ 0 \textit{ otherwise} \end{cases} \tag{11.5}$$

Replacing the hypothetical response pattern with the real scored data, the above function can be viewed as a likelihood function to be maximized by a given set of item parameters. Once items are calibrated in this manner, a likelihood function for the proficiency θ_k is induced from student responses to the calibrated items. This likelihood function for the proficiency θ_k is called the posterior distribution of the θ's for each student.

Proficiency Estimation Using Plausible Values

Most cognitive skills testing is concerned with accurately assessing the performance of individual students 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—that is, 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. Since the uncertainty associated with each θ in such tests is negligible, the distribution of θ, or the joint distribution of θ with other variables, can be approximated using each individual's estimated θ.

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 (Martin, Mullis, & Foy, 2015). This design solicits relatively few responses from each sampled student while maintaining a wide range of content representation when responses are aggregated across all students. With this approach, however, the advantage of estimating population characteristics more efficiently is offset by the inability to make precise statements about individuals. Indeed, the uncertainty associated with individual θ estimates becomes too large to be ignored. In this situation, aggregations of individual student scores can lead to seriously biased estimates of population characteristics (Wingersky, Kaplan, & Beaton, 1987).

Plausible values methodology was developed as a way to address this issue. Instead of first computing estimates of individual θ's and then aggregating these to estimate population parameters, the plausible values approach uses all available data, students' responses to the items they were administered together with all background data, to estimate directly the characteristics of student populations and subpopulations. Although these directly estimated population characteristics could be used for reporting purposes, instead the usual plausible values approach is to generate multiple imputed scores, called plausible values, from the estimated ability distributions and to use these in analyses and reporting, making use of standard statistical software. By including all available background data in the model, a process known as "conditioning," relationships between these background variables and the estimated proficiencies will be appropriately accounted for in the plausible values. Because of this, analyses conducted using plausible values will provide an accurate representation of these underlying relationships. A detailed review of the plausible values methodology is given in Mislevy (1991).[3]

The following is a brief overview of the plausible values approach. Let y represent the responses of all sampled students to background questions or background data of sampled students collected from other sources, and let θ represent the proficiency of interest. If θ were known for all sampled students, it would be possible to compute a statistic $t(\theta,y)$, such as a sample mean or sample percentile point, to estimate a corresponding population quantity T.

Because of the latent nature of the proficiency, however, θ values are not known even for sampled students. The solution to this problem is to follow Rubin (1987) by considering θ as "missing data" and approximate $t(\theta,y)$ by its expectation given (x,y), the data that actually were observed, as follows:

$$\begin{aligned} t^*(x, y) &= E\left| t\left(\underline{\theta}, \underline{y}\right) | \underline{x}, \underline{y} \right| \\ &= \int t\left(\underline{\theta}, \underline{y}\right) p\left(\underline{\theta} \middle| \underline{x}, \underline{y}\right) d\underline{\theta} \end{aligned} \tag{11.6}$$

3 Along with theoretical justifications, Mislevy presents comparisons with standard procedures; discusses biases that arise in some secondary analyses; and offers numerical examples.

It is possible to approximate t^* using random draws from the conditional distribution of the scale proficiencies given the student's item responses x_j, the student's background variables y_j, and model parameters for the items. These values are referred to as imputations in the sampling literature, and as plausible values in large-scale surveys such as PIRLS, TIMSS, NAEP, NALS, and IALLS. The value of θ for any student that would enter into the computation of t is thus replaced by a randomly selected value from his or her conditional distribution. Rubin (1987) proposed repeating this process several times so that the uncertainly associated with imputation can be quantified. For example, the average of multiple estimates of t, each computed from a different set of plausible values, is a numerical approximation of t^* of the above equation; the variance among them reflects the uncertainty due to not observing $\underline{\theta}$. It should be noted that this variance does not include the variability of sampling from the population. That variability is estimated separately by a jackknife variance estimation procedure.

Plausible values are not intended to be estimates of individual student scores, but rather are imputed scores for like students—students with similar response patterns and background characteristics in the sampled population—that may be used to estimate population characteristics correctly. When the underlying model is correctly specified, plausible values will provide consistent estimates of population characteristics, even though they are generally biased estimates of the proficiencies of the individuals with whom they are associated. Taking the average of the plausible values still will not yield suitable estimates of individual student scores.[4]

Plausible values for each student j are drawn from the conditional distribution $P\left(\theta_j \middle| x_j, y_j, \Gamma, \Sigma\right)$, where Γ is a matrix of regression coefficients for the background variables, and Σ is a common variance matrix of residuals. Using standard rules of probability, the conditional probability of proficiency can be represented as:

$$P\left(\theta_j \middle| x_j, y_j, \Gamma, \Sigma\right) \propto P\left(x_j \middle| \theta_j, y_j, \Gamma, \Sigma\right) P\left(\theta_j \middle| y_j, \Gamma, \Sigma\right) = P\left(x_j \middle| \theta_j\right) P\left(\theta_j \middle| y_j, \Gamma, \Sigma\right) \quad (11.7)$$

where θ_j is a vector of scale values, $P(x_j|\theta_j)$ is the product over the scales of the independent likelihoods induced by responses to items within each scale, and $P(\theta_j|y_j, \Gamma, \Sigma)$ is the multivariate joint density of proficiencies for the scales, conditional on the observed values y_j of background responses and parameters Γ and Σ. Item parameter estimates are fixed and regarded as population values in the computations described in this section.

4 For further discussion, see Mislevy, Beaton, Kaplan, and Sheehan (1992).

Conditioning

A multivariate normal distribution was assumed for $P(\theta_j | y_j, \Gamma, \Sigma)$, with a common variance Σ, and with a mean given by a linear model with regression parameters Γ. Since in large-scale studies like PIRLS there are many hundreds of background variables, it is customary to conduct a principal components analysis to reduce the number of variables to be used in Γ. Typically, components accounting for 90 percent of the variance in the data are selected. These principal components are referred to as the conditioning variables and denoted as y^c. The following model is then fit to the data:

$$\theta = \Gamma' y^c + \varepsilon \tag{11.8}$$

where ε is normally distributed with mean zero and variance Σ. As in a regression analysis, Γ is a matrix each of whose columns is the effects for each scale and Σ is the matrix of residual variance between scales.

Note that in order to be strictly correct for all functions Γ of θ, it is necessary that $P(\theta | y)$ be correctly specified for all background variables in the survey. Estimates of functions Γ involving background variables not conditioned in this manner are subject to estimation error due to misspecification. The nature of these errors is discussed in detail in Mislevy (1991). In PIRLS, however, the principal components account for almost all of the variance in the student background variables, so that the computation of marginal means and percentile points of θ for these variables is nearly optimal.

The basic method for estimating Γ and Σ with the Expectation and Maximization (EM) procedure is described in Mislevy (1985) for a single scale case. The EM algorithm requires the computation of the mean θ, and variance Σ, of the posterior distribution in Equation (11.7).

Generating Proficiency Scores

After completing the EM algorithm, plausible values for all sampled students are drawn from the joint distribution of the values of Γ in a three-step process. First, a value of Γ is drawn from a normal approximation to $P(\Gamma, \Sigma | x_j, y_j)$ that fixes Σ at the value $\hat{\Sigma}$ (Thomas, 1993). Second, conditional on the generated value of Γ (and the fixed value of $\Sigma = \hat{\Sigma}$), the mean θ_j and variance Σ_j^p of the posterior distribution in Equation (11.7), where p is the number of scales, are computed using the methods applied in the EM algorithm. In the third step, the proficiency values are drawn independently from a multivariate normal distribution with mean θ_j and variance Σ_j^p. These three steps are repeated five times, producing five imputations of θ_j for each sampled student.

For students with an insufficient number of responses, the Γ's and Σ's described in the previous paragraph are fixed. Hence, all students—regardless of the number of items attempted—are assigned a set of plausible values.

The plausible values can then be employed to evaluate Equation (11.6) for an arbitrary function T as follows:

- Using the first vector of plausible values for each student, evaluate T as if the plausible values were the true values of θ. Denote the result as T_1
- Evaluate the sampling variance of T_1, or Var_1, with respect to students' first vector of plausible values
- Carry out steps 1 and 2 for the second through fifth vectors of plausible values, thus obtaining T_u and Var_u, for $u = 2, \ldots, 5$
- The best estimate of T obtainable from the plausible values is the average of the five values obtained from the different sets of plausible values:

$$\hat{T} = \frac{\sum_u T_u}{5} \qquad (11.9)$$

- An estimate of the variance of $\hat{T}$ is the sum of two components: an estimate of Var_u obtained by averaging as in the previous step, and the variance among the T_u's

Let $\overline{U} = \frac{\sum_u Var_u}{M}$, and let $B_M = \frac{\sum_u \left(T_u - \hat{T}\right)^2}{M-1}$ be the variance among the M plausible values

Then the estimate of the total variance of $\hat{T}$ is:

$$Var\left(\hat{T}\right) = \overline{U} + \left(1 + M^{-1}\right) B_M \qquad (11.10)$$

The first component in $Var\left(\hat{T}\right)$ reflects the uncertainty due to sampling students from the population; the second reflects the uncertainty due to the fact that sampled students' θ's are not known precisely, but only indirectly through x and y.

Working with Plausible Values

The plausible values methodology is used in PIRLS to ensure the accuracy of estimates of the proficiency distributions for the PIRLS populations as a whole and particularly for comparisons between subpopulations. A further advantage of this method is that the variation between the five plausible values generated for each student reflects the uncertainty associated with proficiency estimates for individual students. However, retaining this component of uncertainty requires that additional analytical procedures be used to estimate students' proficiencies.

If the θ values were observed for all sampled students, the statistic $(t-T)/U^{\frac{1}{2}}$ would follow a t-distribution with d degrees of freedom. Then the incomplete-data statistic $(T-\hat{T})/\left[Var(\hat{T})\right]^{\frac{1}{2}}$ is approximately t-distributed, with degrees of freedom (Johnson & Rust, 1992) given by:

$$v = \frac{1}{\frac{f_M^{\ 2}}{M-1} + \frac{(1-f_M)^2}{d}} \tag{11.11}$$

where d is the degrees of freedom for the complete-data statistic, and f_M is the proportion of total variance due to not observing the values:

$$f_M = \frac{(1+M^{-1})\, B_M}{Var(\hat{T})} \tag{11.12}$$

When B_M is small relative to $\overline{U}$, the reference distribution for the incomplete-data statistic differs little from the reference distribution for the corresponding complete-data statistic. If, in addition, d is large, the normal approximation can be used instead of the t-distribution.

For a k-dimensional function T, such as the k coefficients in a multiple regression analysis, each U and $\overline{U}$ is a covariance matrix, and B_M is an average of squares and cross-products rather than simply an average of squares. In this case, the quantity $(\underline{T}-\hat{\underline{T}})\ Var^{-1}(\hat{\underline{T}})\ (\underline{T}-\hat{\underline{T}})'$ is approximately F-distributed with degrees of freedom equal to k and v, with v defined as above but with a matrix generalization of f_M:

$$f_M = (1+M^{-1})\ \mathrm{Trace}\left[B_M\ Var^{-1}(\hat{T})\right]/k \tag{11.13}$$

For the same reason that the normal distribution can approximate the t-distribution, a chi-square distribution with k degrees of freedom can be used in place of the F-distribution for evaluating the significance of the above quantity $(\underline{T}-\hat{\underline{T}})\ Var^{-1}(\hat{\underline{T}})\ (\underline{T}-\hat{\underline{T}})'$.

Statistics $\hat{T}$, the estimates of proficiency conditional on responses to cognitive items and background variables, are consistent estimates of the corresponding population values T, as long as background variables are included in the conditioning variables. The consequences of violating this restriction are described by Beaton and Johnson (1992), Mislevy (1991), and Mislevy and Sheehan (1987). To avoid such biases, the PIRLS analyses include nearly all student background variables, in the form of principal components, as well as the class means to preserve between-class differences—the between-classroom and within-classroom variance structure essential for hierarchical modeling.

References

Beaton, A.E., & Johnson, E.G. (1992). Overview of the scaling methodology used in the National Assessment. *Journal of Educational Measurement, 26*(2), 163–175.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-479). Reading, MA: Addison-Wesley Publishing.

Johnson, E.G., & Rust, K.F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics*, *17*(2), 175–190.

Lord, F.M., & Novick, M.R. (Eds.), (1968). *Statistical theories of mental test scores*. Redding, MA: Addison-Wesley.

Lord, F.M. (1980). *Applications of items response theory to practical testing problems*. Hillsdales, NJ: Lawrence Erlbaum Associates.

Martin, M.O., Mullis, I.V.S., & Foy, P. (2015). Assessment design for PIRLS, PIRLS Literacy, and ePIRLS in 2016. In I.V.S. Mullis & M.O. Martin (Eds.), *PIRLS 2016 Assessment Framework, 2nd Edition* (pp. 55-69). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Mislevy, R.J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association*, *80*, 993–997.

Mislevy, R.J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, *56*, 177–196.

Mislevy, R.J., Beaton, A.E., Kaplan, B., & Sheehan, K.M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*(2), 133–161.

Mislevy, R.J., Johnson, E.G. & Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics*, *17*(2), 131–154.

Mislevy, R.J., & Sheehan, K.M. (1987). Marginal estimation procedures. In A.E. Beaton (Ed.), *Implementing the new design: The NAEP 1983-84 technical report* (No. 15-TR-20) (pp. 293–360). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*(2), 159–176.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Thomas, N. (1993). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics*, *2*, 309–322.

Van Der Linden, W.J., & Hambleton, R. (1996). *Handbook of Modern Item Response Theory*. New York. Springer-Verlag.

von Davier, M. (2014). Imputing proficiency data under planned missingness in population models. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis* (pp. 175–201). Boca Raton: Chapman & Hall/CRC.

von Davier, M., & Sinharay, S. (2014). Analytics in international large-scale assessments: item response theory and population models. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis* (pp. 155–174). Boca Raton: Chapman & Hall/CRC.

Wingersky, M., Kaplan, B., & Beaton, A.E. (1987). Joint estimation procedures. In A.E. Beaton (Ed.), *Implementing the new design: The NAEP 1983-84 technical report* (No. 15-TR-20) (pp.285–92). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Yamamoto, K., & Kulick, E. (2000). Scaling methodology and procedures for the TIMSS mathematics and science scales. In M.O. Martin, K.D. Gregory, & S.E. Stemler (Eds.), *TIMSS 1999 technical report*. Chestnut Hill, MA: Boston College.

CHAPTER 12

Scaling the PIRLS 2016 Achievement Data

Pierre Foy
Liqun Yin

Overview

The PIRLS 2016 assessment had ambitious goals for broad coverage of the reading purposes and processes, as described in its assessment framework, and for measuring trends across assessment cycles. Given this broad coverage, PIRLS used a matrix-sampling booklet design such that each student was administered only a subset of the entire PIRLS item pool (see Chapter 4 of *PIRLS 2016 Assessment Framework, 2nd Edition*). Given the complexities of the data collection and the need to have student scores on the entirety of each assessment for analysis and reporting purposes, PIRLS relied on item response theory (IRT) scaling to describe student achievement and to provide accurate measures of trends. As each student responded to only a part of the assessment item pool, the PIRLS scaling approach used multiple imputation—or plausible values—methodology to obtain proficiency scores in reading for all students. To enhance the reliability of the student scores, the PIRLS scaling approach uses conditioning, a process in which student responses to the items are combined with information about students' backgrounds.

This scaling chapter begins with a general description of the PIRLS scaling approach and its use of plausible values. It then describes the concurrent calibration method used specifically to measure trends. Next, it explains how the proficiency scores are generated through the use of conditioning and describes the process of transforming the proficiency scores to place them on the metric used to measure trends. A special section then describe how the PIRLS Literacy 2016 achievement data were scaled and placed on the PIRLS reading reporting scale and another section describes the scaling of the ePIRLS 2016 achievement data. A theoretical description of the PIRLS scaling methodology can be found in Chapter 11: PIRLS 2016 Achievement Scaling Methodology.

Implementing the PIRLS Scaling Procedure

The application of IRT scaling and plausible values methodology to the data from the PIRLS 2016 assessment involved four major tasks: calibrating the achievement items (estimating model parameters for each item), creating principal components from the student and parent questionnaire data for use in conditioning, generating reading proficiency scores, and placing these proficiency scores on the metric used to report trend results from previous assessments. New for PIRLS 2016, the PIRLS Literacy achievement results were reported on the PIRLS reading scale. Also, in order to report trends back to its predecessor assessment, prePIRLS 2011, the 2011 scores were re-calibrated. The scaling procedure also generated proficiency scores for the domains of overall reading: the purposes for reading and the processes of comprehension.

Linking Assessments Cycles with Concurrent Calibration

The metric of the PIRLS reporting scale was originally established in PIRLS 2001 by setting the mean of the national average scores for all countries that participated in PIRLS 2001 to 500 and the standard deviation to 100. To enable measurement of trends over time, achievement data from successive PIRLS assessments were transformed to this same metric. This is done by concurrently scaling the data from each successive assessment with the data from the previous assessment—a process known as concurrent calibration—and applying linear transformations to place the results from each successive assessment on the same scale as the results from the previous assessment. This procedure enables PIRLS to measure trends across all four assessment cycles: 2001, 2006, 2011, and 2016.[1]

The first step in linking the assessments for trend scaling is to estimate (calibrate) the item parameters for the items in the current assessment through a concurrent calibration of the data from the current assessment and from the previous assessment. In 2016, the PIRLS concurrent calibration consisted of combining achievement data from the 2016 and 2011 assessments.

In linking successive assessments, concurrent calibration relies on having a large proportion of trend items, items that are retained from one assessment to the next. The PIRLS 2016 assessment consisted of 6 literary passages with their items and 6 informational passages with their items. In PIRLS 2016, 3 of the literary passages and 3 of the informational passages consisted of newly developed items. The remaining 3 literary passages and 3 informational passages were carried forward from the PIRLS 2011 assessment and are the basis for linking PIRLS 2016 to the PIRLS achievement scale and maintaining trends over time. Exhibit 12.1 lists the number of items present for the PIRLS 2016 concurrent calibration by item type and by purposes for reading and processes of comprehension.

1 See Mazzeo and von Davier (2014) for a discussion of the linking procedure used by PIRLS.

Exhibit 12.1: PIRLS 2016 Reading Items for Concurrent Calibration

Item Type	Points	Items Released in 2011		Items Common in 2011 and 2016		Items Introduced in 2016		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple-Choice	1	29	29	45	45	41	41	115	115
Constructed Response	1	12	12	16	16	33	33	61	61
	2	12	24	15	30	17	34	44	88
	3	1	3	5	15	3	9	9	27
Total		**54**	**68**	**81**	**106**	**94**	**117**	**229**	**291**

PIRLS 2016 Reading Items for Concurrent Calibration by Reading Purposes and Comprehension Processes

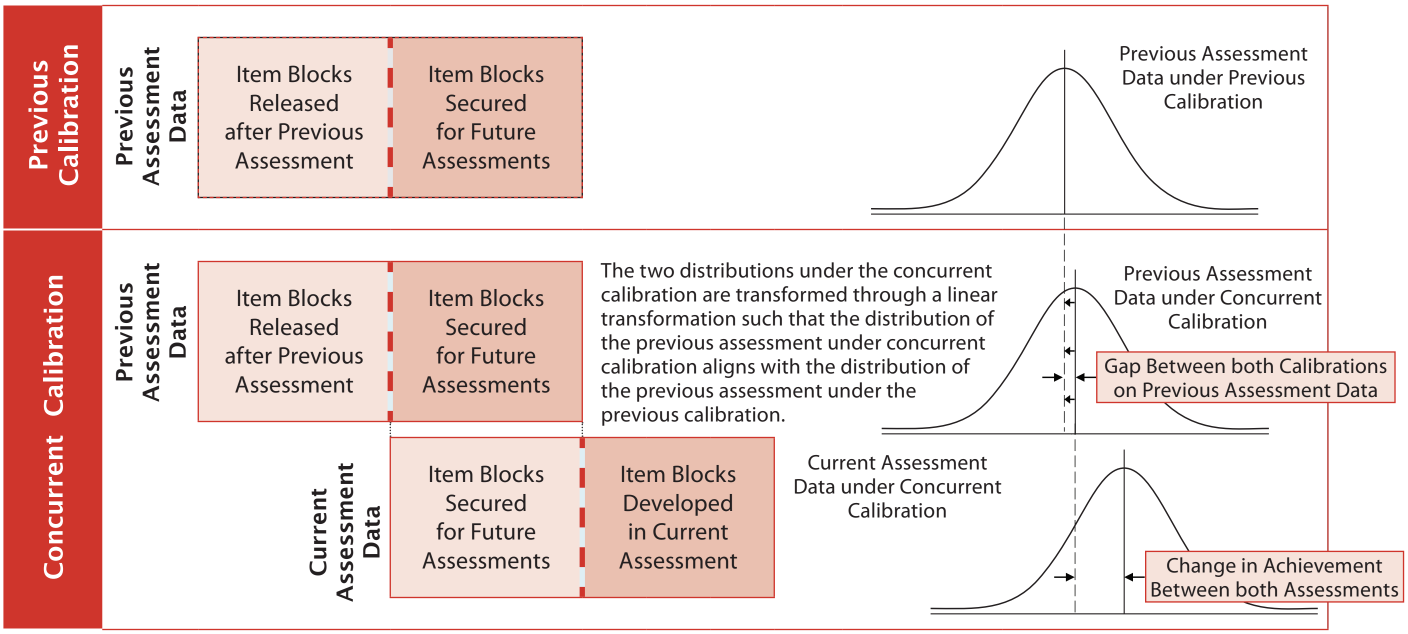
Purposes for Reading Items	Items Released in 2011		Items Common in 2011 and 2016		Items Introduced in 2016		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Literary Experience	28	35	44	55	46	58	118	148
Acquire and Use Information	26	33	37	51	48	59	111	143
Processes of Comprehension	**Items Released in 2011**		**Items Common in 2011 and 2016**		**Items Introduced in 2016**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**	**Items**	**Points**
Retrieving and Straightforward Inferencing	33	36	46	51	57	65	136	152
Interpreting, Integrating, and Evaluating	21	32	35	55	37	52	93	139
Total	**54**	**68**	**81**	**106**	**94**	**117**	**229**	**291**

In concurrent calibration, item parameters for the current assessment are estimated based on the data from both the current and previous assessments, recognizing that some items (the trend items) are common to both. It is then possible to estimate the latent ability distributions of students in both assessments using the item parameters from the concurrent calibration. The difference between these two distributions is the change in achievement between the previous and current assessments.

After the concurrent calibration, the next step is to make use of student and parent context data from their respective questionnaires, in a process called conditioning, to enhance the reliability of the proficiency scores estimated. Once these proficiency scores are estimated, the next step consists of finding a linear transformation that transforms the proficiency distribution of the previous assessment data under the concurrent calibration to match the proficiency distribution of these same data under the calibration that was done in the previous assessment. The final step entails applying this linear transformation to the current assessment data scaled using the concurrent calibration. This places the current assessment data on the trend scale.

Exhibit 12.2 illustrates how the concurrent calibration approach is applied in the context of PIRLS trend scaling. The gap between the distributions of the previous assessment data under the previous calibration and under the concurrent calibration is typically small and is the result of slight differences in the item parameter estimates from the two calibrations (Exhibit 12.2, second panel). The linear transformation removes this gap by shifting the two distributions from the concurrent calibration such that the distribution of the previous assessment data from the concurrent calibration aligns with the distribution of the previous assessment data from the previous calibration,[2] while preserving the gap between the previous and current assessment data under the concurrent calibration. This latter gap is the change in achievement between the previous and current assessments that PIRLS sets out to measure as trend.

Exhibit 12.2: Concurrent Calibration Model Used for PIRLS



Calibrating the PIRLS 2016 Assessment Data

Item calibration was conducted by the TIMSS & PIRLS International Study Center using the commercially-available Parscale software (Muraki & Bock, 1991) and included data from the previous assessment (PIRLS 2011) and data from the 2016 assessment for countries that participated

2 The difference between the ability distributions of the previous assessment data under the two calibrations is a measure of the linkage error in the trend scaling procedure.

in both assessment cycles. The calibration used all available item response data from each country's student samples and from both current and previous assessments. All student samples were weighted so that each country contributed equally to the item calibration. Exhibit 12.3 shows the sample sizes for scaling the PIRLS 2016 data. A total of 40 countries from PIRLS 2016 contributed to the concurrent calibration. Norway's data at the fourth grade were included in the concurrent calibration.

Exhibit 12.3: Sample Sizes for PIRLS 2016 Achievement Scales

Country	Concurrent Calibration		Proficiency Estimation	
	2016	2011	2016	2011
Australia	6,341	6,126	6,341	6,126
Austria	4,360	4,670	4,360	4,670
Azerbaijan	4,990	4,881	5,994	4,881
Bahrain	—	—	5,480	—
Belgium (Flemish)	—	—	5,198	—
Belgium (French)	4,623	3,727	4,623	3,727
Bulgaria	4,281	5,261	4,281	5,261
Canada	18,245	23,206	18,245	23,206
Chile	—	—	4,294	—
Chinese Taipei	4,326	4,293	4,326	4,293
Czech Republic	5,537	4,556	5,537	4,556
Denmark	3,508	4,594	3,508	4,594
England	5,095	3,927	5,095	3,927
Finland	4,896	4,640	4,896	4,640
France	4,767	4,438	4,767	4,438
Georgia	5,741	4,796	5,741	4,796
Germany	3,959	4,000	3,959	4,000
Hong Kong SAR	3,349	3,875	3,349	3,875
Hungary	4,623	5,204	4,623	5,204
Iran, Islamic Rep. of	4,385	5,758	4,385	5,758
Ireland	4,607	4,524	4,607	4,524
Israel	4,041	4,186	4,041	4,186
Italy	3,940	4,189	3,940	4,189
Kazakhstan	—	—	4,925	—
Latvia	—	—	4,157	—
Lithuania	2,947	4,661	4,317	4,661
Macao SAR	—	—	4,059	—

Exhibit 12.3: Sample Sizes for PIRLS 2016 Achievement Scales (Continued)

Country	Concurrent Calibration		Proficiency Estimation	
	2016	2011	2016	2011
Malta	3,647	3,548	3,647	3,548
Morocco	5,489	7,805	5,489	7,805
Netherlands	4,206	3,995	4,206	3,995
New Zealand	5,646	5,644	5,646	5,644
Northern Ireland	3,693	3,586	3,693	3,586
Norway (5)	—	—	4,232	—
Oman	9,234	10,394	9,234	10,394
Poland	—	—	4,413	—
Portugal	4,642	4,085	4,642	4,085
Qatar	9,077	4,120	9,077	4,120
Russian Federation	4,577	4,461	4,577	4,461
Saudi Arabia	4,741	4,507	4,741	4,507
Singapore	6,488	6,367	6,488	6,367
Slovak Republic	5,451	5,630	5,451	5,630
Slovenia	4,499	4,512	4,499	4,512
Spain	14,595	8,580	14,595	8,580
Sweden	4,525	4,622	4,525	4,622
Trinidad and Tobago	4,177	3,948	4,177	3,948
United Arab Emirates	16,471	14,618	16,471	14,618
United States	4,425	12,726	4,425	12,726
Benchmarking Participants				
Buenos Aires, Argentina	—	—	4,382	—
Ontario, Canada	—	—	4,270	4,561
Quebec, Canada	—	—	3,179	4,244
Norway (4)	4,354	3,190	4,354	3,190
Moscow City, Russian Fed.	—	—	4,289	—
Eng/Afr/Zulu - RSA (5)	—	—	5,282	—
Andalusia, Spain	—	—	4,169	4,333
Madrid, Spain	—	—	3,794	—
Abu Dhabi, UAE	—	—	4,188	4,146
Dubai, UAE	—	—	7,859	6,061
Total	**228,498**	**231,850**	**309,042**	**255,195**

The item parameters estimated from these concurrent calibrations, based on the countries that have participated in both the previous and current assessments, were used to estimate student proficiency for all countries and benchmarking entities participating in the PIRLS 2016 assessment. These item parameters also were used to estimate student proficiency in the purposes for reading and processes of comprehension domains. Student proficiency was estimated for a total of 47 countries and 10 benchmarking participants, as shown in Exhibit 12.3. The item parameters estimated from the PIRLS 2016 concurrent calibration are presented in Appendix 13A.

Treatment of Omitted and Not-Reached Responses

Given the matrix-sampling design used by PIRLS, whereby a student is administered only a sample of the assessment items (from one literary passage and one informational passage) most items are missing by design for each student. However, missing data can also result from a student not answering an item, which can occur when the student does not know the answer, omits the item by mistake, or does not have sufficient time to attempt the item. An item is considered "not reached" when—within part 1 or part 2 of a booklet[3]—the item itself and the item immediately preceding it are not answered, and there are no other items completed in the remainder of that part of the booklet.

Not-reached items are treated differently in estimating item parameters and in generating student proficiency scores. In estimating the values of the item parameters, items in the assessment booklets that are considered not to have been reached by students are treated as if they have not been administered. This approach is considered optimal for parameter estimation. However, not-reached items are considered as incorrect responses when student proficiency scores are generated.

Evaluating Fit of IRT Models to the PIRLS Assessment Data

After the item calibration was completed, checks were performed to verify that the item parameters obtained from Parscale adequately reproduced the observed distribution of student responses across the proficiency continuum. The fit of the IRT models to the PIRLS assessment data was examined by comparing the item response function curves generated using the item parameters estimated from the data with the empirical item response functions calculated from the latent abilities estimated for each student that responded to the item. When the empirical results for an item fall near the fitted curves, the IRT model fits the data well and provides an accurate and reliable measurement of the underlying proficiency scale. Graphical plots of these response function curves are called item characteristic curves (ICC).

The plots in the Exhibits 12.4 and 12.5 show examples of the empirical and fitted item response functions for dichotomously scored (right/wrong) multiple-choice and constructed response items,

3 The PIRLS assessment consist of two parts, with a break in between.

respectively. In each plot, the horizontal axis represents the proficiency scale, and the vertical axis represents the probability of a correct response. The fitted curve based on the estimated item parameters is shown as a solid line. Empirical results are represented by circles. The empirical results are obtained by first dividing the proficiency scale into intervals of equal size and then counting the number of students responding to the item whose estimated latent abilities (EAP scores) from Parscale fall in each interval. Then the proportion of students in each interval that responded correctly to the item is calculated. In the exhibits, the center of each circle represents this empirical proportion of correct responses. The size of each circle is proportional to the number of students contributing to the estimation of the empirical proportion correct.

Exhibit 12.4: Example of Item Response Function for a Dichotomous Multiple-Choice Item from the PIRLS 2016 Assessment

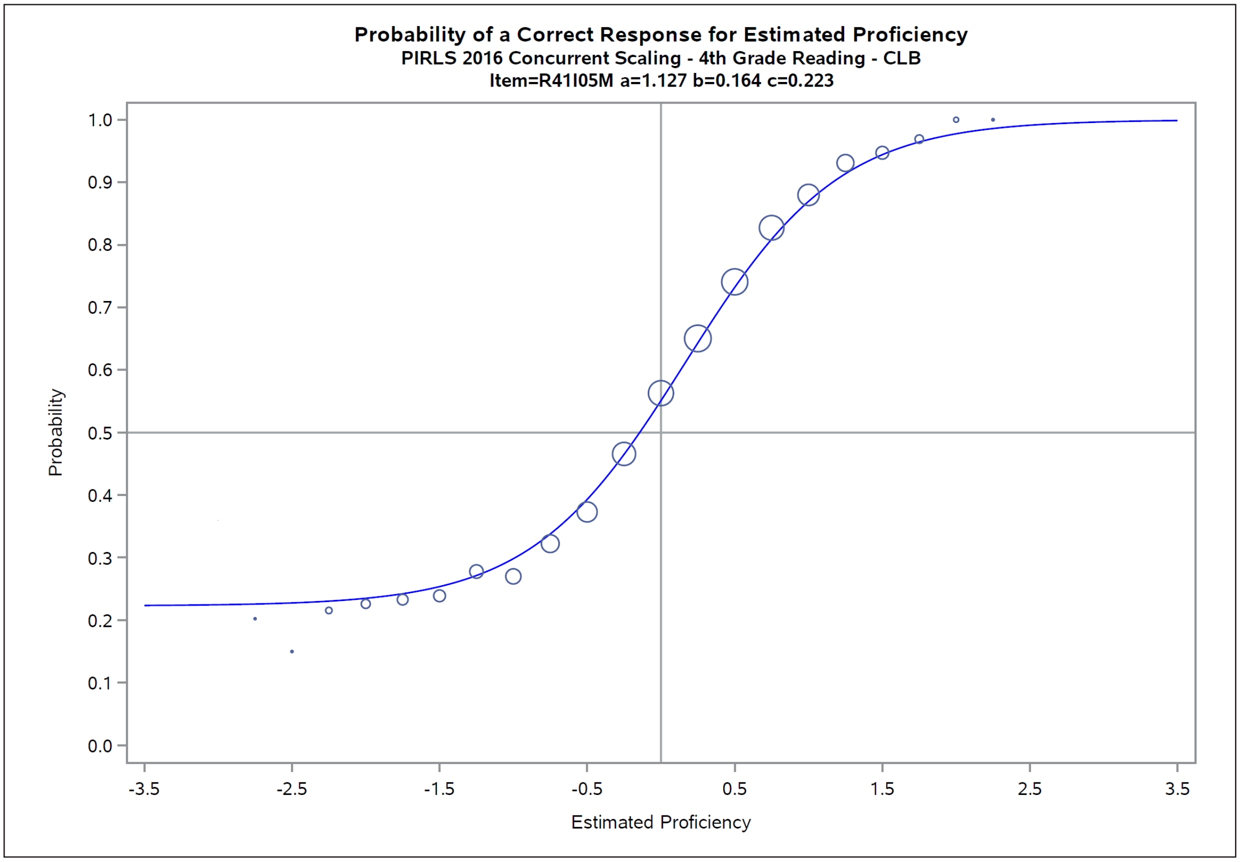
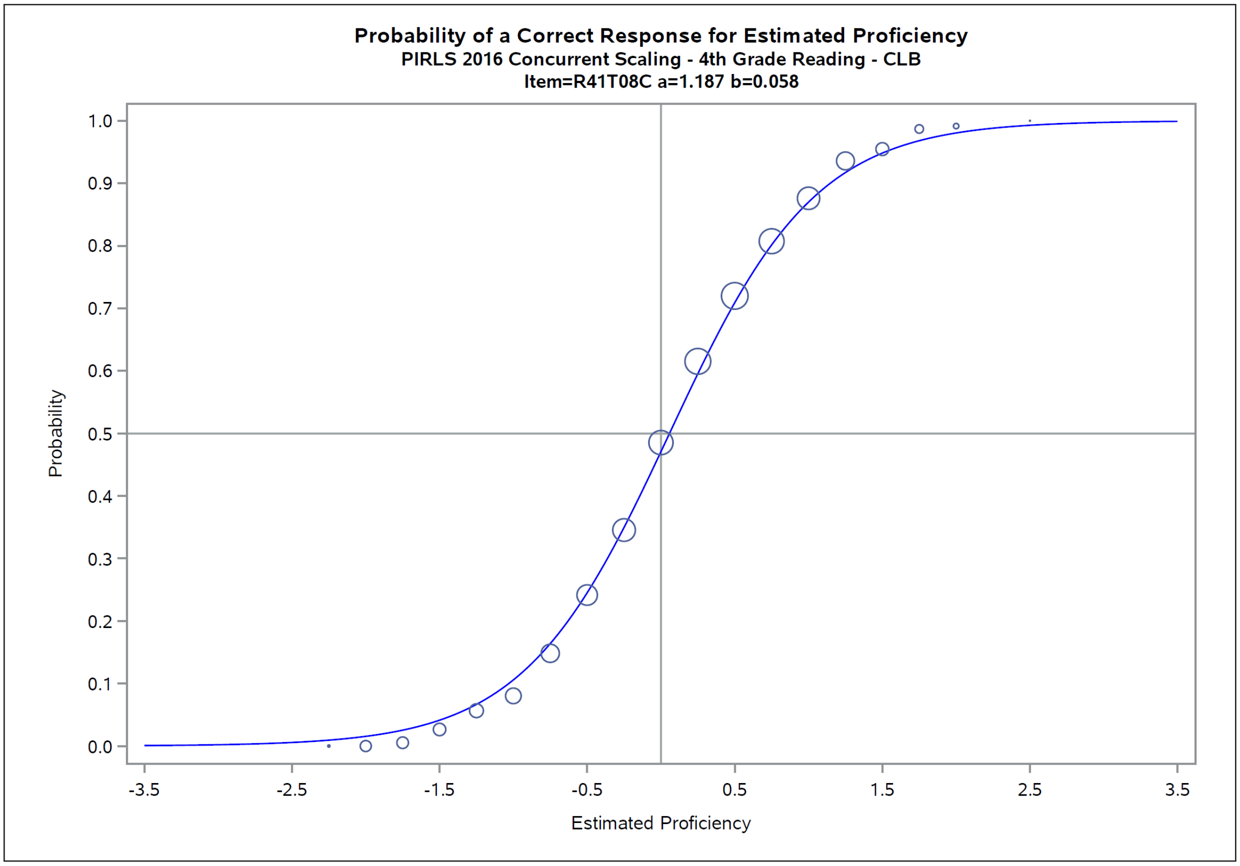
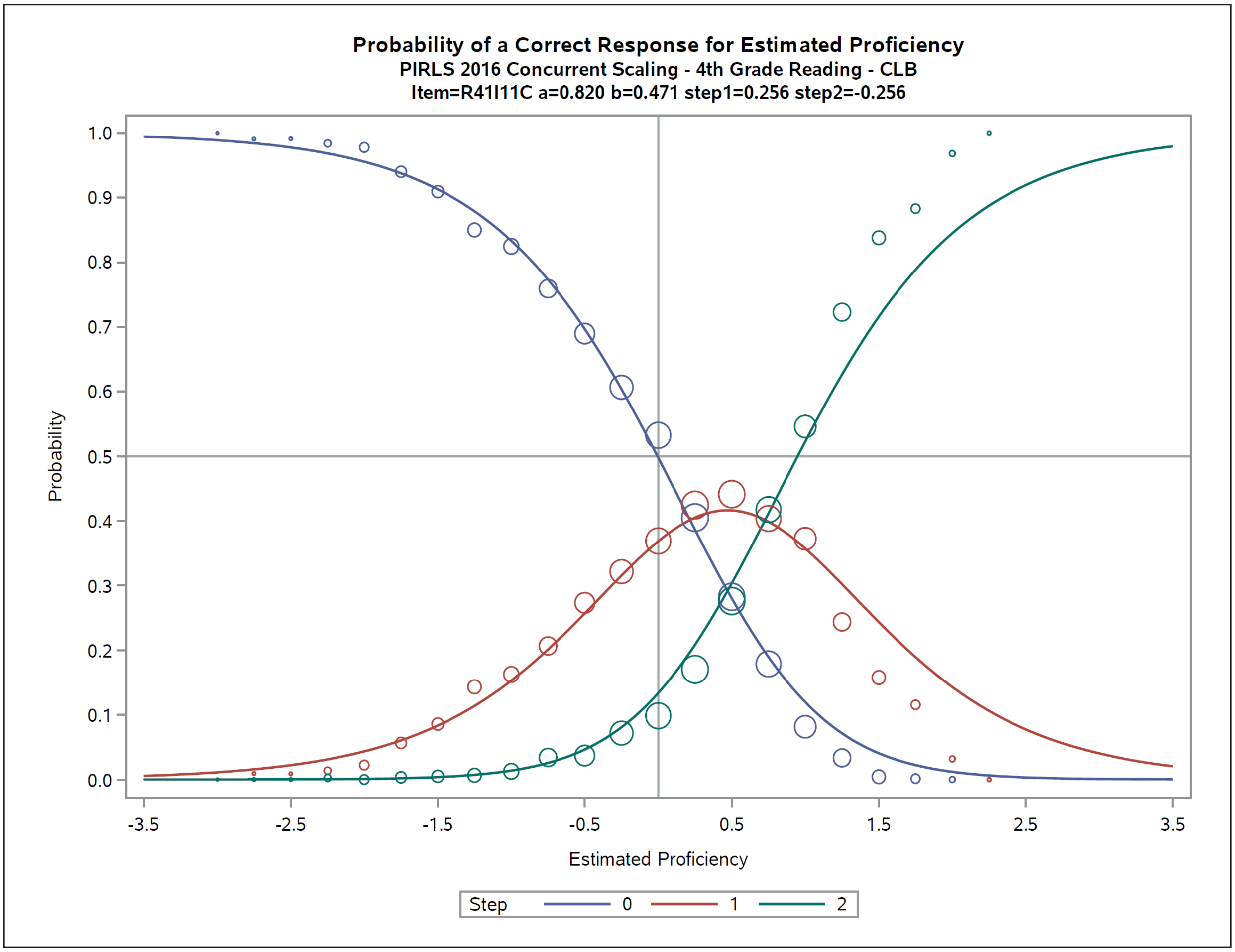


Exhibit 12.5: Example of Item Response Function for a Dichotomous Constructed Response Item from the PIRLS 2016 Assessment



The plot in Exhibit 12.6 shows the empirical and fitted item response functions for a polytomous item (scored 0, 1, or 2). As for the dichotomous item plots, the horizontal axis represents the proficiency scale, but in this example the vertical axis represents the probability of having a response in a given response category. The fitted curves based on the estimated item parameters are shown as solid lines and again the empirical results are represented by circles. The interpretation of the circles is the same as in Exhibits 12.4 and 12.5. The curve starting at the top left of the chart plots the probability of a score of zero on the item. This probability decreases as proficiency increases. The bell-shaped curve shows the probability of a score of one point—partial credit, starting low for low-ability students, reaching a maximum for medium-ability students, and decreasing for high-ability students. The curve ending at the top right corner of the chart shows the probability of a score of two points—full credit, starting low for low-ability students and increasing as proficiency increases.

Exhibit 12.6: Example of Item Response Function for a Polytomous Constructed Response Item from the PIRLS 2016 Assessment



Variables for Conditioning the PIRLS Assessment Data

Conditioning is the practice of using all available students' context information to improve the reliability of the estimated student proficiency scores. Ideally, all context data would be included in the conditioning model, but because PIRLS has so many student context variables that could be used in conditioning, the TIMSS & PIRLS International Study Center follows the practice established by NAEP and followed by other large-scale studies of using principal components analysis to reduce the number of variables while explaining most of their common variance. Principal components for the PIRLS student context variables (including parent context variables) were constructed as follows:

- For categorical variables (questions with a small number of fixed response options), a dummy coded variable was created for each response option, with a value of one if the option is chosen and zero otherwise. If a student omitted or was not administered a particular question, all dummy coded variables associated with that question were assigned the value zero.

- Background variables with numerous response options (such as year of birth) were recoded using criterion scaling.[4] This was done by replacing the response option with the mean interim achievement score of all students choosing that option. Criterion scaling maximizes the correlation between the scaled variable and achievement. For PIRLS, the interim achievement score was the reading EAP scores produced from the item calibrations.
- Separately for each country, all the dummy-coded and criterion-scaled variables were included in a principal components analysis. Those principal components accounting for up to 90 percent of the variance of the context variables were retained for use as conditioning variables.[5] Because the principal components analysis was performed separately for each country, different numbers of principal components were required to account for 90% of the common variance in each country's context variables.

In addition to the principal components, student gender (dummy coded), the language of the test (dummy coded), an indicator of the classroom in the school to which a student belongs (criterion scaled), and an optional country-specific variable (dummy coded) were included as primary conditioning variables, thereby accounting for most of the variance between students and preserving the between-classroom and within-classroom variance structure in the scaling model. Exhibit 12.7 provides details on the conditioning models used for proficiency estimation in PIRLS 2016.

4 The process of generating criterion-scaled variables is described in Beaton (1969).

5 The number of principal components retained is limited to no more than 5% of a country's student sample size, thereby possibly reducing the percentage of variance accounted for, to avoid over-specification of the conditioning model.

Exhibit 12.7: Conditioning Models for PIRLS 2016 Achievement Scales

Country	2016				2011			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Australia	2	539	278	90	2	545	286	90
Austria	2	544	218	79	2	543	233	80
Azerbaijan	3	533	299	89	2	546	244	80
Bahrain	3	545	274	85	—	—	—	—
Belgium (Flemish)	2	541	259	85	—	—	—	—
Belgium (French)	2	526	231	82	2	548	186	73
Bulgaria	2	529	214	81	2	527	263	88
Canada	5	521	293	90	6	540	305	90
Chile	2	518	214	79	—	—	—	—
Chinese Taipei	2	542	216	79	2	553	214	77
Czech Republic	2	536	276	88	2	551	227	80
Denmark	2	545	175	71	2	555	229	78
England	2	248	145	90	2	243	137	90
Finland	3	544	244	84	3	550	232	81
France	2	543	238	82	2	547	221	78
Georgia	3	545	287	88	2	551	239	80
Germany	2	541	197	77	2	552	200	75
Hong Kong SAR	2	541	167	71	2	555	193	73
Hungary	2	521	231	82	2	539	260	85
Iran, Islamic Rep. of	2	545	219	80	2	555	287	86
Ireland	2	545	230	82	2	549	226	80
Israel	3	507	202	78	3	525	209	78
Italy	2	539	197	74	3	551	209	75
Kazakhstan	3	527	246	82	—	—	—	—
Latvia	3	545	207	77	—	—	—	—
Lithuania	4	524	215	79	2	547	233	80
Macao SAR	4	545	202	75	—	—	—	—
Malta	2	537	182	71	2	555	177	69
Morocco	2	545	274	86	2	549	324	90
Netherlands	2	539	210	81	2	546	199	76
New Zealand	7	515	278	90	8	549	282	88
Northern Ireland	2	507	184	79	2	544	179	75
Norway (5)	3	526	211	78	—	—	—	—

Exhibit 12.7: Conditioning Models for PIRLS 2016 Achievement Scales (Continued)

Country	2016				2011			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Oman	3	545	317	90	3	553	323	90
Poland	2	532	220	81	—	—	—	—
Portugal	2	544	232	81	2	542	204	77
Qatar	3	542	307	90	3	544	206	75
Russian Federation	2	521	228	81	2	527	223	80
Saudi Arabia	3	545	237	80	3	544	225	78
Singapore	2	545	300	90	2	555	307	90
Slovak Republic	3	545	272	87	3	547	281	88
Slovenia	2	540	224	81	2	547	225	80
Spain	8	538	303	90	7	544	304	90
Sweden	2	521	226	82	2	547	231	81
Trinidad and Tobago	2	513	208	77	2	525	197	75
United Arab Emirates	6	545	316	90	5	541	317	90
United States	9	250	147	90	9	237	141	90
Benchmarking Participants								
Buenos Aires, Argentina	2	530	219	80	—	—	—	—
Ontario, Canada	3	519	213	81	3	540	228	82
Quebec, Canada	3	519	158	70	3	540	212	78
Norway (4)	3	527	217	79	3	553	159	69
Moscow City, Russian Fed.	2	521	214	78	—	—	—	—
Eng/Afr/Zulu - RSA (5)	4	539	264	85	—	—	—	—
Andalusia, Spain	2	538	208	77	2	541	216	77
Madrid, Spain	2	537	189	73	—	—	—	—
Abu Dhabi, UAE	3	545	209	76	3	541	207	75
Dubai, UAE	4	545	306	90	3	541	303	90

Generating IRT Proficiency Scores for the PIRLS Assessment Data

Educational Testing Service's DGROUP program (Rogers, Tang, Lin, & Kandathil, 2006) was used to generate the IRT proficiency scores. This program takes as input the students' responses to the items they were given, the item parameters estimated at the calibration stage, and the conditioning variables, and generates as output the plausible values that represent student proficiency.

A useful feature of DGROUP is its ability to perform multi-dimensional scaling using the responses to all items across the proficiency scales and the correlations among the scales to improve the reliability of each individual scale. The multi-dimensional scaling feature of DGROUP also was used to generate proficiency scores for the PIRLS 2016 domains. The estimation of proficiency scores for the purposes for reading and the processes of comprehension relied on multidimensional IRT models using the item parameters estimated for the overall reading scale as well the same conditioning variables. PIRLS 2016 used two two-dimensional scaling models, one to estimate proficiency scores for the two purposes for reading and a second for the two processes of comprehension.

In addition to generating plausible values for the overall reading scale from the 2016 assessment data, the item parameters estimated at the calibration stage also were used to generate plausible values for the PIRLS 2011 assessment for the countries included in the concurrent calibration. These additional plausible values were used to establish the linear transformation necessary to place the 2016 assessment data on the PIRLS reading trend scale.

Transforming the Overall Scores to Measure Trends

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. These linear transformations were given by:

$$PV^{*}_{k,i} = A_{k,i} + B_{k,i} \times PV_{k,i} \tag{12.1}$$

where

$PV_{k,i}$ is the PIRLS 2016 plausible value i of scale k prior to transformation;

$PV^{*}_{k,i}$ is the PIRLS 2016 plausible value i of scale k after transformation; and

$A_{k,i}$ and $B_{k,i}$ are the linear transformation constants.

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. From these calculations, the linear transformation constants were defined as:

$$B_{k,i} = \sigma_{k,i} / \sigma^*_{k,i} \tag{12.2}$$

$$A_{k,i} = \mu_{k,i} - B_{k,i} \cdot \mu^*_{k,i} \tag{12.3}$$

where

$\mu_{k,i}$ is the international mean of scale k based on plausible value i published in 2011;

$\mu^*_{k,i}$ is the international mean of scale k based on plausible value i from the 2011 assessment based on the 2016 concurrent calibration;

$\sigma_{k,i}$ is the international standard deviation of scale k based on plausible value i published in 2011;

$\sigma^*_{k,i}$ is the international standard deviation of scale k based on plausible value i from the 2011 assessment based on the 2016 concurrent calibration.

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. Exhibit 12.8 shows the PIRLS 2016 transformation constants for overall reading.

Exhibit 12.8: Linear Transformation Constants for PIRLS 2016 Achievement Scales

Overall Reading	PIRLS 2011 Published Scores		PIRLS 2011 Re-scaled Scores		$A_{k,i}$	$B_{k,i}$
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	514.88796	93.40789	-0.02153	0.96698	516.96808	96.59763
PV2	514.33588	94.16192	-0.01873	0.96533	516.16294	97.54392
PV3	514.10484	93.95296	-0.01702	0.96329	515.76531	97.53376
PV4	514.09822	94.15851	-0.01852	0.96502	515.90514	97.57133
PV5	514.19052	93.93593	-0.01874	0.96576	516.01365	97.26663

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.

The linear transformation constants for overall reading also were applied to the scales for the purposes for reading and the processes of comprehension. In this approach to measuring trends in the purposes and processes, achievement changes over time are established in the context of achievement in overall reading. Trends are not established separately for each purpose or process; rather differential changes in performance in the domains are considered in the context of trends in overall reading.

Scaling the PIRLS Literacy 2016 Achievement Data

Launched in 2011 as prePIRLS, PIRLS Literacy 2016 is a reading assessment intended for populations of readers that would find the PIRLS reading assessment too challenging. Although a less demanding assessment, PIRLS Literacy was designed to allow the reading achievement of participating countries to be reported on the PIRLS reading trend scale. To that end, PIRLS and PIRLS Literacy in 2016 shared four passages to establish a psychometric link between the two assessments. Two shared passages were PIRLS passages with their usual structure of a text accompanied by a set of items related to that text. Two shared passages were PIRLS Literacy passages with their items interspersed within the accompanying text.

Exhibit 12.9 shows the number of items present in the PIRLS Literacy 2016 assessment by item type and domain. There was a total of 183 items in the PIRLS Literacy assessment, 59 of them shared with the PIRLS reading assessment.

Exhibit 12.9: PIRLS Literacy 2016 Items for Calibration

Item Type	Points	PIRLS Literacy Shared Items		PIRLS Literacy Unique Items		Total	
		Items	Points	Items	Points	Items	Points
Multiple-Choice	1	29	29	61	61	90	90
Constructed Response	1	18	18	49	49	67	67
	2	11	22	12	24	23	46
	3	1	3	2	6	3	9
Total		**59**	**72**	**124**	**140**	**183**	**212**

PIRLS Literacy 2016 Items for Calibration by Reading Purposes and Comprehension Processes

Purposes for Reading	PIRLS Literacy Shared Items		PIRLS Literacy Unique Items		Total	
	Items	Points	Items	Points	Items	Points
Literary Experience	30	36	63	71	93	107
Acquire and Use Information	29	36	61	69	90	105
Processes of Comprehension	**PIRLS Literacy Shared Items**		**PIRLS Literacy Unique Items**		**Total**	
	Items	**Points**	**Items**	**Points**	**Items**	**Points**
Retrieving and Straightforward Inferencing	40	44	94	101	134	145
Interpreting, Integrating, and Evaluating	19	28	30	39	49	67
Total	**59**	**72**	**124**	**140**	**183**	**212**

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. Exhibit 12.10 shows the sample sizes for scaling the PIRLS Literacy data. A total of six countries participated and all were included in the item calibration—including data from Denmark's benchmarking participation in PIRLS Literacy 2016 at the 3rd grade.

Exhibit 12.10: Sample Sizes for PIRLS Literacy 2016 Achievement Scales

Country	Item Calibration	Proficiency Estimation
Egypt	6,957	6,957
Iran, Islamic Rep. of	4,381	4,381
Kuwait	4,609	4,609
Morocco	5,453	5,453
South Africa	12,810	12,810
Benchmarking Participants		
Denmark (3)	3,600	3,600
Total	**37,810**	**37,810**

The item calibration step was based on a straightforward calibration of the PIRLS Literacy 2016 achievement items from the six participating countries. The item parameters for the PIRLS Literacy items were placed on the PIRLS reading metric by fixing the parameters of the items in the four shared passages to the values estimated from the PIRLS 2016 concurrent calibration. The item parameters estimated from the PIRLS Literacy 2016 item calibration are presented in Appendix 12B. The 59 link items, whose item parameters were fixed, are marked with asterisks.

The conditioning for PIRLS Literacy 2016 was done in exactly the same way as for PIRLS, as was the estimation of proficiency scores using the DGROUP software. This included overall reading scores for the PIRLS Literacy countries and scores for the PIRLS purposes for reading and processes of comprehension. Exhibit 12.11 provides details on the conditioning models used for the PIRLS Literacy 2016 proficiency estimation.

Exhibit 12.11: PIRLS Literacy 2016 Conditioning Models for Proficiency Estimation

Country	2016			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Egypt	2	545	304	90
Iran, Islamic Rep. of	2	545	219	80
Kuwait	3	535	230	80
Morocco	2	545	272	85
South Africa	12	539	323	90
Benchmarking Participants				
Denmark (3)	2	545	180	72

The final step in the process consisted of placing students' performance on the PIRLS Literacy 2016 assessment on the PIRLS reading reporting scale. This was done by applying the appropriate linear transformation to the estimated proficiency scores. The PIRLS Literacy 2016 item calibration resulted in item parameters on the same metric as the PIRLS 2016 concurrent calibration—by fixing the parameters of the 59 link items. Thus, placing the PIRLS Literacy 2016 achievement scores on the PIRLS reporting scale was accomplished by using the PIRLS 2016 reading linear transformation constants, as presented in Exhibit 12.8. These linear transformation constants were applied to the PIRLS Literacy 2016 overall reading achievement scores, as well as the achievement scores on the purposes for reading and the processes of comprehension.

In 2011, PIRLS Literacy's predecessor prePIRLS was reported as its own scale, although its item parameters were estimated on the same item parameter metric, capitalizing on Colombia's participation in both PIRLS and prePIRLS in 2011. However, with South Africa having participated in both prePIRLS in 2011 and PIRLS Literacy in 2016, there was a need to place their 2011 results on the PIRLS trend scale. To that end, it was necessary to re-transform their achievement scores—overall reading, as well as the purposes and processes—using the PIRLS 2011 linear transformation constants, as given in Exhibit 15 of the Scaling the TIMSS and PIRLS Achievement 2011 Data chapter of Methods and Procedures in TIMSS and PIRLS 2011.

Scaling the ePIRLS 2016 Achievement Data

ePIRLS 2016 is a new computer-based assessment of online informational reading, consisting of five tasks, designed to assess fourth grade students' ability to use the internet in a school context. With ePIRLS designed as an extension of PIRLS reading, students participating in ePIRLS 2016 were expected also to participate in PIRLS 2016. Thus, 14 countries and two benchmarking entities participated in both PIRLS and ePIRLS in 2016. Exhibit 12.12 lists the ePIRLS participants and their effective sample sizes across both PIRLS and ePIRLS assessments in 2016.

Exhibit 12.12: PIRLS 2016 and ePIRLS 2016 Sample Sizes

Country	PIRLS Sample Size	ePIRLS Sample Size	Percentage Overlap	ePIRLS Students not in PIRLS	Percentage not in PIRLS
Canada	18,245	8,871	48.6	261	2.9
Chinese Taipei	4,326	4,299	99.4	32	0.7
Denmark	3,508	2,506	71.4	120	4.6
Georgia	5,741	5,557	96.8	100	1.8
Ireland	4,607	2,473	53.7	82	3.2
Israel	4,041	3,798	94.0	135	3.4
Italy	3,940	3,767	95.6	95	2.5
Norway (5)	4,232	3,610	85.3	104	2.8
Portugal	4,642	4,558	98.2	78	1.7
Singapore	6,488	6,320	97.4	100	1.6
Slovenia	4,499	4,303	95.6	67	1.5
Sweden	4,525	3,879	85.7	109	2.7
United Arab Emirates	16,471	15,566	94.5	441	2.8
United States	4,425	4,090	92.4	16	0.4
Total	**89,690**	**73,597**	**82.1**	**1,740**	**1.9**
Benchmarking Participants					
Abu Dhabi, UAE	4,188	3,980	95.0	86	2.1
Dubai, UAE	7,859	7,471	95.1	155	2.0

In general, ePIRLS 2016 participants were successful in having nearly all their sampled students participate in both assessments, with a few notable exceptions. In Canada, some provinces did not take part in ePIRLS and a subsample of Quebec's PIRLS schools participated in ePIRLS. In Ireland, because of limitations in the number of computers available in many schools, random subsamples of PIRLS students participated in ePIRLS. In Denmark, Norway, and Sweden, some PIRLS schools were unable to participate in ePIRLS, generally arising from the absence of compatible computers for the ePIRLS assessment. It is worth pointing out that a small proportion of students—less than 2% internationally—took part in the ePIRLS assessment, but not in the PIRLS assessment. These students were removed from the ePIRLS samples. Thus, only students that participated in both PIRLS and ePIRLS assessments were retained in the ePIRLS samples.

Exhibit 12.13 shows the number of items present in the ePIRLS 2016 assessment by item type and process of comprehension. The exhibit also includes the PIRLS 2016 items since they were included in the ePIRLS item calibration. There was a total of 91 items in the ePIRLS 2016 assessment. The 175 PIRLS 2016 items were also included in the item calibration, with fixed item parameters from the PIRLS 2016 concurrent calibration.

Exhibit 12.13: ePIRLS 2016 Items for Calibration

Item Type	Points	ePIRLS 2016 Items		PIRLS 2016 Items		Total	
		Items	Points	Items	Points	Items	Points
Multiple-Choice	1	36	36	86	86	122	122
Constructed Response	1	37	37	49	49	86	86
	2	15	30	32	64	47	94
	3	3	9	8	24	11	33
Total		**91**	**112**	**175**	**223**	**266**	**335**

ePIRLS 2016 Items by Comprehension Process

Processes of Comprehension	ePIRLS 2016 Items		PIRLS 2016 Items		Total	
	Items	Points	Items	Points	Items	Points
Retrieving and Straightforward Inferencing	49	54	103	116	152	170
Interpreting, Integrating, and Evaluating	42	58	72	107	114	165
Total	**91**	**112**	**175**	**223**	**266**	**335**

The ePIRLS scaling methodology adopted the same four steps of calibration, conditioning, generating proficiency scores, and placing those scores on the PIRLS reading scale. All 14 ePIRLS countries were included in the item calibration, including their responses to the PIRLS and ePIRLS items. The item parameters for the ePIRLS 2016 items were placed on the PIRLS reading metric by fixing the parameters of the PIRLS 2016 items to the values estimated from the PIRLS 2016 concurrent calibration. The item parameters estimated from the ePIRLS 2016 item calibration are presented in Appendix 12C. Although the PIRLS 2016 items were included in the ePIRLS item calibration, they are not included in Appendix 12C as they are in every way identical to the parameters estimated for PIRLS 2016 and presented in Appendix 12A.

Exhibit 12.14 provides details on the conditioning models used for the ePIRLS 2016 proficiency estimation. Although ePIRLS used the same set of conditioning variables from the PIRLS student and parents questionnaires, the resulting conditioning matrices were not necessarily identical to PIRLS since the ePIRLS samples sizes were not the same as the PIRLS sample sizes. The DGROUP software was used to estimate ePIRLS proficiency scores, including overall ePIRLS online informational reading scores and scores for the two PIRLS processes of comprehension.

Exhibit 12.14: ePIRLS 2016 Conditioning Models for Proficiency Estimation

Country	2016			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Canada	5	521	279	90
Chinese Taipei	2	542	214	79
Denmark	2	545	125	62
Georgia	3	545	277	87
Ireland	2	545	123	62
Israel	3	507	189	76
Italy	2	539	188	73
Norway (5)	3	526	180	73
Portugal	2	544	227	80
Singapore	2	545	299	90
Slovenia	2	540	215	80
Sweden	2	521	193	77
United Arab Emirates	6	545	315	90
United States	9	250	147	90
Benchmarking Participants				
Abu Dhabi, UAE	3	545	199	75
Dubai, UAE	4	545	306	90

The final step in the process consisted of placing students' performance on the ePIRLS 2016 assessment on the PIRLS reading reporting scale. This was done by applying the appropriate linear transformation to the estimated proficiency scores. The ePIRLS 2016 item calibration resulted in item parameters on the same metric as the PIRLS reading metric—by fixing the parameters of all PIRLS 2016 items. Thus, placing the ePIRLS achievement scores on the PIRLS reporting scale was accomplished by using the PIRLS 2016 reading linear transformation constants, as presented in Exhibit 12.8. These linear transformation constants were applied to the ePIRLS 2016 overall online informational reading achievement scores, as well as the achievement scores on the two processes of comprehension.

References

Beaton, A.E. (1969). Criterion scaling of questionnaire items. *Socio-Economic Planning Sciences*, 2, 355–362.

Mazzeo, J., and von Davier, M. (2014) Linking scales in International Large-Scale Assessments. In L. Rutkowski, M. von Davier, & D. Rutkowski, *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis* (pp. 229-258). Boca Raton: Chapman & Hall/CRC.

Muraki, E., & Bock, R.D. (1991). *PARSCALE: Parameter scaling of rating data* [computer software]. Chicago, IL: Scientific Software, Inc.

Rogers, A., Tang, C., Lin, M.-J., & Kandathil, M. (2006). DGROUP [computer software]. Princeton, NJ: Educational Testing Service.

Appendix 12A: PIRLS 2016 Item Parameters from Concurrent Calibration

Item Parameters from PIRLS 2016 Concurrent Item Calibration

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
Items Released in 2011:						
R21E01M	1.375 (0.091)	-1.108 (0.077)	0.378 (0.035)			
R21E02M	1.143 (0.071)	-0.344 (0.061)	0.287 (0.027)			
R21E03M	0.552 (0.045)	-0.337 (0.127)	0.153 (0.039)			
R21E04M	1.452 (0.091)	-1.076 (0.067)	0.330 (0.033)			
R21E05C	0.619 (0.017)	-0.495 (0.025)		-0.539 (0.051)	0.539 (0.044)	
R21E06M	1.323 (0.076)	-0.228 (0.046)	0.242 (0.023)			
R21E07C	0.657 (0.021)	-0.243 (0.023)		0.180 (0.042)	-0.180 (0.036)	
R21E08M	1.321 (0.081)	0.477 (0.032)	0.162 (0.015)			
R21E09C	0.534 (0.021)	0.735 (0.031)		0.534 (0.042)	-0.534 (0.053)	
R21E10C	0.964 (0.035)	-0.174 (0.025)				
R21E11M	0.901 (0.071)	0.266 (0.065)	0.249 (0.026)			
R21E12C	0.780 (0.026)	0.173 (0.019)		0.315 (0.032)	-0.315 (0.032)	
R31P01M	1.106 (0.057)	-0.484 (0.051)	0.149 (0.025)			
R31P02C	0.856 (0.032)	-0.776 (0.035)				
R31P03C	1.095 (0.038)	-0.529 (0.025)				
R31P04M	0.990 (0.067)	0.583 (0.040)	0.130 (0.017)			
R31P05C	0.649 (0.019)	0.421 (0.020)		-0.310 (0.040)	0.310 (0.042)	
R31P06C	1.353 (0.046)	-0.613 (0.023)				
R31P07C	0.941 (0.024)	-0.117 (0.016)		-0.137 (0.031)	0.137 (0.028)	
R31P08M	1.090 (0.068)	-0.335 (0.063)	0.263 (0.028)			
R31P09C	1.199 (0.041)	-0.485 (0.024)				
R31P10M	1.769 (0.095)	-0.520 (0.039)	0.240 (0.023)			
R31P11M	1.152 (0.063)	-0.226 (0.048)	0.171 (0.023)			
R31P12M	1.342 (0.076)	0.133 (0.036)	0.182 (0.019)			
R31P13M	1.325 (0.072)	-0.753 (0.056)	0.221 (0.030)			
R31P14C	1.184 (0.041)	0.028 (0.020)				
R31P15C	0.630 (0.023)	0.397 (0.022)		0.173 (0.038)	-0.173 (0.041)	
R31P16C	0.783 (0.036)	0.744 (0.033)				
R21N01M	0.852 (0.059)	-0.640 (0.099)	0.281 (0.037)			
R21N02C	0.780 (0.030)	-0.494 (0.033)				

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R21N03C	0.747 (0.028)	1.061 (0.029)		0.353 (0.031)	-0.353 (0.047)	
R21N04M	1.175 (0.074)	0.207 (0.042)	0.207 (0.020)			
R21N05M	1.610 (0.093)	-0.856 (0.051)	0.276 (0.029)			
R21N06M	1.457 (0.076)	-0.475 (0.042)	0.201 (0.023)			
R21N07M	1.074 (0.066)	-0.043 (0.052)	0.205 (0.024)			
R21N08C	0.933 (0.034)	-0.269 (0.026)				
R21N09M	1.178 (0.072)	-0.337 (0.058)	0.270 (0.027)			
R21N10M	0.878 (0.073)	0.249 (0.073)	0.284 (0.028)			
R21N11C	0.555 (0.016)	-0.010 (0.023)		-0.555 (0.050)	0.555 (0.048)	
R21N12C	0.636 (0.022)	0.115 (0.021)		0.080 (0.041)	-0.080 (0.039)	
R31G01M	1.116 (0.066)	-0.513 (0.064)	0.257 (0.030)			
R31G02C	0.680 (0.028)	-0.160 (0.031)				
R31G03M	1.100 (0.067)	-0.303 (0.060)	0.253 (0.027)			
R31G04C	0.863 (0.038)	0.982 (0.036)				
R31G05M	1.178 (0.091)	0.481 (0.045)	0.288 (0.020)			
R31G06M	1.019 (0.059)	-0.309 (0.059)	0.195 (0.027)			
R31G07M	1.101 (0.066)	0.088 (0.045)	0.180 (0.021)			
R31G08CZ	0.792 (0.028)	0.977 (0.026)		0.264 (0.029)	-0.264 (0.042)	
R31G09M	0.877 (0.061)	0.079 (0.064)	0.197 (0.026)			
R31G10C	0.993 (0.038)	0.566 (0.024)				
R31G11M	1.612 (0.107)	0.336 (0.034)	0.302 (0.017)			
R31G12C	0.465 (0.018)	1.639 (0.059)		-0.863 (0.063)	0.863 (0.086)	
R31G13CZ	0.819 (0.019)	0.157 (0.013)		-0.280 (0.039)	0.108 (0.044)	0.171 (0.036)
R31G14M	1.312 (0.088)	0.359 (0.039)	0.241 (0.019)			
Items Common in 2011 and 2016:						
R11F01M	1.334 (0.049)	-0.627 (0.034)	0.148 (0.018)			
R11F02M	0.666 (0.038)	-0.848 (0.111)	0.243 (0.037)			
R11F03M	0.920 (0.039)	-0.666 (0.054)	0.157 (0.024)			
R11F04M	1.307 (0.053)	-0.831 (0.044)	0.228 (0.023)			
R11F05M	0.940 (0.045)	-0.255 (0.052)	0.217 (0.022)			
R11F06C	0.776 (0.023)	-0.152 (0.021)				
R11F07C	0.503 (0.010)	0.375 (0.018)		-0.896 (0.041)	0.896 (0.043)	
R11F08C	1.149 (0.029)	-0.328 (0.017)				
R11F09C	1.011 (0.022)	-0.627 (0.015)		0.074 (0.027)	-0.074 (0.020)	

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R11F10C	0.846 (0.026)	-1.419 (0.039)				
R11F11M	0.739 (0.045)	0.208 (0.061)	0.192 (0.023)			
R11F12C	0.618 (0.014)	0.642 (0.017)		-0.471 (0.032)	0.471 (0.036)	
R11F13M	1.124 (0.054)	-0.199 (0.046)	0.270 (0.021)			
R21Y01M	1.097 (0.055)	0.154 (0.038)	0.253 (0.017)			
R21Y02M	1.649 (0.070)	-0.204 (0.029)	0.288 (0.016)			
R21Y03C	0.815 (0.025)	0.564 (0.021)				
R21Y04M	1.273 (0.056)	0.093 (0.031)	0.222 (0.015)			
R21Y05M	1.721 (0.070)	0.086 (0.022)	0.226 (0.013)			
R21Y06M	1.533 (0.062)	0.042 (0.025)	0.209 (0.014)			
R21Y07M	0.792 (0.037)	-1.011 (0.079)	0.182 (0.030)			
R21Y08M	1.360 (0.058)	-0.271 (0.035)	0.261 (0.018)			
R21Y09C	0.956 (0.020)	-0.551 (0.015)		0.078 (0.027)	-0.078 (0.020)	
R21Y10C	0.749 (0.024)	0.574 (0.023)				
R21Y11M	1.411 (0.065)	0.035 (0.032)	0.284 (0.016)			
R21Y12C	0.706 (0.012)	-0.001 (0.014)		-1.154 (0.039)	1.154 (0.038)	
R21Y13C	0.760 (0.017)	0.378 (0.011)		0.594 (0.026)	-0.219 (0.028)	-0.375 (0.031)
R21Y14C	0.576 (0.013)	0.222 (0.016)		-0.549 (0.036)	0.549 (0.036)	
R31M01M	1.451 (0.062)	-0.877 (0.043)	0.268 (0.023)			
R31M02C	1.218 (0.033)	-0.957 (0.023)				
R31M03M	1.330 (0.057)	-0.004 (0.031)	0.228 (0.016)			
R31M04C	0.560 (0.020)	0.357 (0.028)				
R31M05M	1.551 (0.083)	0.112 (0.034)	0.424 (0.015)			
R31M06M	1.159 (0.063)	0.368 (0.036)	0.283 (0.016)			
R31M07M	1.619 (0.068)	-0.350 (0.031)	0.276 (0.017)			
R31M08M	1.382 (0.058)	-0.498 (0.038)	0.265 (0.020)			
R31M09C	0.759 (0.015)	-0.030 (0.017)		1.143 (0.027)	-1.143 (0.024)	
R31M10C	0.623 (0.021)	0.413 (0.025)				
R31M11M	0.854 (0.043)	-0.692 (0.075)	0.262 (0.029)			
R31M12M	1.196 (0.050)	0.106 (0.030)	0.162 (0.015)			
R31M13M	2.100 (0.089)	-0.642 (0.027)	0.256 (0.018)			
R31M14M	2.283 (0.087)	-0.195 (0.019)	0.197 (0.013)			
R31M15M	1.298 (0.057)	0.065 (0.031)	0.218 (0.016)			
R31M16C	1.207 (0.031)	0.058 (0.015)				
R31M17CZ	0.612 (0.014)	0.003 (0.013)		0.071 (0.038)	0.228 (0.037)	-0.299 (0.032)

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R11L01M	0.532 (0.027)	-2.275 (0.169)	0.146 (0.053)			
R11L02M	0.749 (0.056)	0.685 (0.054)	0.241 (0.020)			
R11L03C	0.616 (0.020)	-0.474 (0.029)				
R11L04C	0.667 (0.012)	0.418 (0.015)		1.643 (0.030)	-1.051 (0.032)	-0.592 (0.043)
R11L05M	1.186 (0.057)	0.352 (0.030)	0.206 (0.014)			
R11L06C	0.656 (0.021)	0.191 (0.023)				
R11L07M	0.772 (0.045)	0.474 (0.045)	0.154 (0.018)			
R11L08C	0.801 (0.019)	0.612 (0.015)		0.703 (0.021)	-0.703 (0.027)	
R11L09M	0.963 (0.043)	-0.809 (0.061)	0.226 (0.026)			
R11L10C	0.732 (0.019)	0.681 (0.016)		0.231 (0.024)	-0.231 (0.029)	
R11L11M	0.912 (0.042)	-0.354 (0.052)	0.189 (0.022)			
R11L12C	0.735 (0.017)	0.509 (0.016)		0.810 (0.023)	-0.810 (0.028)	
R21K01C	0.422 (0.013)	-0.891 (0.033)		0.186 (0.052)	-0.186 (0.039)	
R21K02C	0.807 (0.023)	-0.559 (0.025)				
R21K03M	1.004 (0.047)	0.081 (0.039)	0.184 (0.018)			
R21K04M	1.062 (0.096)	0.979 (0.045)	0.391 (0.014)			
R21K05C	0.969 (0.026)	0.137 (0.017)				
R21K06M	1.489 (0.067)	0.052 (0.029)	0.281 (0.015)			
R21K07C	0.682 (0.016)	0.143 (0.015)		0.119 (0.027)	-0.119 (0.027)	
R21K08M	0.994 (0.052)	0.354 (0.037)	0.197 (0.017)			
R21K09M	1.203 (0.056)	-0.010 (0.037)	0.246 (0.018)			
R21K10C	0.785 (0.017)	0.778 (0.015)		-0.397 (0.027)	0.397 (0.031)	
R21K11M	1.070 (0.056)	0.260 (0.039)	0.240 (0.017)			
R21K12C	0.576 (0.013)	-0.110 (0.014)		0.446 (0.040)	-0.084 (0.037)	-0.362 (0.034)
R31W01C	0.718 (0.017)	-0.584 (0.018)		0.243 (0.032)	-0.243 (0.024)	
R31W02C	0.800 (0.017)	0.278 (0.013)		-0.107 (0.024)	0.107 (0.025)	
R31W03M	1.347 (0.052)	-0.063 (0.027)	0.162 (0.014)			
R31W04C	0.842 (0.024)	-0.687 (0.026)				
R31W05M	1.264 (0.068)	0.497 (0.030)	0.257 (0.014)			
R31W06M	0.753 (0.034)	-0.999 (0.080)	0.147 (0.031)			
R31W07CZ	0.879 (0.017)	0.509 (0.010)		-0.079 (0.025)	0.169 (0.029)	-0.090 (0.027)
R31W08M	1.355 (0.063)	-0.093 (0.037)	0.307 (0.018)			
R31W09M	0.951 (0.054)	0.565 (0.036)	0.178 (0.016)			
R31W10M	1.289 (0.056)	0.320 (0.026)	0.164 (0.013)			
R31W11C	1.467 (0.038)	0.551 (0.013)				

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R31W12M	1.514 (0.081)	0.706 (0.023)	0.219 (0.011)			
R31W13C	0.862 (0.028)	0.791 (0.023)				
Items Introduced in 2016:						
L21B01C	0.677 (0.036)	-2.530 (0.108)				
L21B02M	0.958 (0.067)	-2.353 (0.150)	0.101 (0.072)			
L21B03M *	0.721 (0.048)	-2.406 (0.130)	0.250 (0.000)			
L21B04C	0.706 (0.032)	-1.260 (0.057)				
L21B05M	0.883 (0.061)	-1.315 (0.127)	0.223 (0.054)			
L21B06M *	0.440 (0.033)	-1.847 (0.144)	0.250 (0.000)			
L21B07C	0.745 (0.033)	-1.160 (0.052)				
L21B08C	0.840 (0.036)	-1.359 (0.053)				
L21B09C	0.855 (0.036)	-1.140 (0.047)				
L21B10M	0.660 (0.057)	-0.338 (0.133)	0.192 (0.046)			
L21B11M	0.979 (0.062)	-1.224 (0.103)	0.186 (0.048)			
L21B12M	0.738 (0.055)	-2.202 (0.212)	0.127 (0.091)			
L21B13C	0.542 (0.022)	-0.967 (0.040)		0.693 (0.067)	-0.693 (0.044)	
L21B14M	1.075 (0.063)	-0.597 (0.066)	0.149 (0.032)			
L21B15C	0.419 (0.026)	0.318 (0.053)				
L21B16C	0.435 (0.019)	-0.185 (0.035)		0.092 (0.065)	-0.092 (0.059)	
L21B17C	0.705 (0.027)	-0.414 (0.027)		0.684 (0.048)	-0.684 (0.035)	
R41H01M	0.947 (0.101)	-0.667 (0.166)	0.582 (0.044)			
R41H02M	1.058 (0.066)	-0.743 (0.079)	0.196 (0.038)			
R41H03C	1.172 (0.044)	0.257 (0.022)				
R41H04C	0.671 (0.043)	1.644 (0.083)				
R41H05M	1.030 (0.082)	0.131 (0.070)	0.286 (0.029)			
R41H06C	0.700 (0.022)	-0.264 (0.024)		-0.089 (0.045)	0.089 (0.039)	
R41H07M	0.895 (0.085)	0.873 (0.053)	0.151 (0.022)			
R41H08C	0.691 (0.042)	1.492 (0.071)				
R41H09M	0.649 (0.064)	0.272 (0.110)	0.166 (0.040)			
R41H10M	1.307 (0.086)	-0.179 (0.056)	0.263 (0.028)			
R41H11M	1.402 (0.094)	-0.537 (0.065)	0.311 (0.034)			
R41H12M	1.350 (0.104)	0.002 (0.061)	0.369 (0.027)			
R41H13C	0.541 (0.018)	0.753 (0.023)		-0.089 (0.054)	0.270 (0.063)	-0.181 (0.068)
R41H14C	0.990 (0.042)	0.307 (0.025)				

* Items with fixed guessing parameters.

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R41H15C	1.250 (0.050)	-0.385 (0.028)				
R41H16C	1.027 (0.046)	0.577 (0.026)				
R41O01M	0.942 (0.065)	-0.993 (0.109)	0.254 (0.046)			
R41O02C	0.859 (0.036)	-0.965 (0.044)				
R41O03C	1.020 (0.046)	0.868 (0.030)				
R41O04C	0.595 (0.024)	1.145 (0.038)		-0.160 (0.046)	0.160 (0.061)	
R41O05C	0.630 (0.020)	0.228 (0.022)		-0.260 (0.046)	0.260 (0.045)	
R41O06M	1.437 (0.097)	-0.212 (0.055)	0.315 (0.028)			
R41O07C	0.629 (0.022)	-0.953 (0.037)		0.036 (0.062)	-0.036 (0.043)	
R41O08C	0.847 (0.036)	-0.064 (0.029)				
R41O09C	1.205 (0.045)	-0.059 (0.023)				
R41O10C	0.815 (0.026)	0.236 (0.019)		-0.006 (0.036)	0.006 (0.035)	
R41O11M	1.365 (0.096)	0.096 (0.050)	0.295 (0.025)			
R41O12M	1.283 (0.074)	-0.341 (0.051)	0.169 (0.027)			
R41O13C	0.567 (0.014)	0.348 (0.018)		-1.197 (0.072)	0.794 (0.079)	0.403 (0.057)
L21E01C	0.796 (0.047)	-3.130 (0.125)				
L21E02C	0.771 (0.038)	-2.116 (0.083)				
L21E03M	1.343 (0.084)	-0.521 (0.059)	0.270 (0.029)			
L21E04M	1.106 (0.074)	-0.378 (0.070)	0.262 (0.031)			
L21E05M	1.118 (0.076)	-1.568 (0.110)	0.242 (0.052)			
L21E06M	0.936 (0.081)	-2.465 (0.221)	0.304 (0.097)			
L21E07C	0.463 (0.018)	-0.810 (0.042)		-0.018 (0.071)	0.018 (0.055)	
L21E08M	1.023 (0.064)	-0.749 (0.081)	0.201 (0.036)			
L21E09M	0.620 (0.085)	0.464 (0.145)	0.340 (0.041)			
L21E10C	0.859 (0.047)	-2.716 (0.101)				
L21E11M	1.205 (0.074)	-0.817 (0.071)	0.226 (0.035)			
L21E12C	1.116 (0.057)	-2.136 (0.066)				
L21E13C	0.528 (0.027)	-0.275 (0.046)				
L21E14C	0.493 (0.020)	0.186 (0.032)		0.852 (0.053)	-0.852 (0.053)	
L21E15C	0.795 (0.040)	-2.219 (0.087)				
L21E16C	0.706 (0.032)	-0.959 (0.051)				
L21E17M	1.047 (0.065)	-0.467 (0.069)	0.171 (0.032)			
R41I01C	0.793 (0.036)	-1.522 (0.063)				
R41I02M	1.034 (0.085)	0.623 (0.048)	0.185 (0.021)			
R41I03C	0.560 (0.022)	0.201 (0.026)		0.330 (0.047)	-0.330 (0.047)	

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R41I04C	0.685 (0.026)	0.470 (0.023)		0.301 (0.038)	-0.301 (0.042)	
R41I05M	1.127 (0.080)	0.164 (0.054)	0.223 (0.025)			
R41I06M	1.260 (0.102)	0.487 (0.047)	0.283 (0.021)			
R41I07C	0.968 (0.030)	0.502 (0.017)		-0.008 (0.030)	0.008 (0.032)	
R41I08M	1.594 (0.099)	-0.579 (0.052)	0.264 (0.030)			
R41I09C	1.152 (0.044)	-0.035 (0.023)				
R41I10M	0.952 (0.079)	0.265 (0.070)	0.244 (0.029)			
R41I11C	0.820 (0.030)	0.471 (0.020)		0.256 (0.033)	-0.256 (0.036)	
R41I12M	1.061 (0.070)	0.155 (0.051)	0.145 (0.025)			
R41I13C	0.804 (0.036)	0.207 (0.030)				
R41I14C	0.801 (0.042)	0.944 (0.040)				
R41I15C	0.755 (0.041)	0.690 (0.037)				
R41T01M	1.085 (0.071)	-0.973 (0.092)	0.242 (0.044)			
R41T02C	0.690 (0.024)	-0.516 (0.028)		0.269 (0.048)	-0.269 (0.037)	
R41T03C	0.943 (0.029)	0.071 (0.017)		0.034 (0.033)	-0.034 (0.030)	
R41T04C	1.239 (0.046)	0.055 (0.021)				
R41T05M	0.756 (0.077)	0.383 (0.095)	0.244 (0.035)			
R41T06C	1.381 (0.052)	-0.579 (0.026)				
R41T07C	0.780 (0.025)	0.622 (0.021)		-0.202 (0.037)	0.202 (0.041)	
R41T08C	1.187 (0.044)	0.058 (0.022)				
R41T09M	1.560 (0.103)	0.500 (0.031)	0.182 (0.016)			
R41T10C	1.264 (0.047)	-0.212 (0.023)				
R41T11C	0.859 (0.022)	0.518 (0.014)		-0.426 (0.043)	0.319 (0.050)	0.106 (0.042)
R41T12M	0.999 (0.086)	-0.050 (0.089)	0.346 (0.035)			
R41T13M	1.111 (0.075)	0.432 (0.041)	0.112 (0.020)			
R41T14C	0.533 (0.030)	0.131 (0.043)				
R41T15M	0.897 (0.091)	0.594 (0.072)	0.235 (0.029)			
R41T16M	1.289 (0.098)	-0.046 (0.064)	0.286 (0.030)			

Appendix 12B: PIRLS Literacy 2016 Item Parameters from Item Calibration

Item Parameters from PIRLS Literacy 2016 Item Calibration

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
Items Shared with PIRLS 2016 (Fixed Item Parameters):						
L21B01C *	0.677 (0.036)	-2.530 (0.108)				
L21B02M *	0.958 (0.067)	-2.353 (0.150)	0.101 (0.072)			
L21B03M *	0.721 (0.048)	-2.406 (0.130)	0.250 (0.000)			
L21B04C *	0.706 (0.032)	-1.260 (0.057)				
L21B05M *	0.883 (0.061)	-1.315 (0.127)	0.223 (0.054)			
L21B06M *	0.440 (0.033)	-1.847 (0.144)	0.250 (0.000)			
L21B07C *	0.745 (0.033)	-1.160 (0.052)				
L21B08C *	0.840 (0.036)	-1.359 (0.053)				
L21B09C *	0.855 (0.036)	-1.140 (0.047)				
L21B10M *	0.660 (0.057)	-0.338 (0.133)	0.192 (0.046)			
L21B11M *	0.979 (0.062)	-1.224 (0.103)	0.186 (0.048)			
L21B12M *	0.738 (0.055)	-2.202 (0.212)	0.127 (0.091)			
L21B13C *	0.542 (0.022)	-0.967 (0.040)		0.693 (0.067)	-0.693 (0.044)	
L21B14M *	1.075 (0.063)	-0.597 (0.066)	0.149 (0.032)			
L21B15C *	0.419 (0.026)	0.318 (0.053)				
L21B16C *	0.435 (0.019)	-0.185 (0.035)		0.092 (0.065)	-0.092 (0.059)	
L21B17C *	0.705 (0.027)	-0.414 (0.027)		0.684 (0.048)	-0.684 (0.035)	
R11F01M *	1.334 (0.049)	-0.627 (0.034)	0.148 (0.018)			
R11F02M *	0.666 (0.038)	-0.848 (0.111)	0.243 (0.037)			
R11F03M *	0.920 (0.039)	-0.666 (0.054)	0.157 (0.024)			
R11F04M *	1.307 (0.053)	-0.831 (0.044)	0.228 (0.023)			
R11F05M *	0.940 (0.045)	-0.255 (0.052)	0.217 (0.022)			
R11F06C *	0.776 (0.023)	-0.152 (0.021)				
R11F07C *	0.503 (0.010)	0.375 (0.018)		-0.896 (0.041)	0.896 (0.043)	
R11F08C *	1.149 (0.029)	-0.328 (0.017)				
R11F09C *	1.011 (0.022)	-0.627 (0.015)		0.074 (0.027)	-0.074 (0.020)	
R11F10C *	0.846 (0.026)	-1.419 (0.039)				
R11F11M *	0.739 (0.045)	0.208 (0.061)	0.192 (0.023)			
R11F12C *	0.618 (0.014)	0.642 (0.017)		-0.471 (0.032)	0.471 (0.036)	
R11F13M *	1.124 (0.054)	-0.199 (0.046)	0.270 (0.021)			

* Items with fixed item parameters estimated in PIRLS 2016 concurrent item calibration.

Item Parameters from PIRLS Literacy 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
L21E01C *	0.796 (0.047)	-3.130 (0.125)				
L21E02C *	0.771 (0.038)	-2.116 (0.083)				
L21E03M *	1.343 (0.084)	-0.521 (0.059)	0.270 (0.029)			
L21E04M *	1.106 (0.074)	-0.378 (0.070)	0.262 (0.031)			
L21E05M *	1.118 (0.076)	-1.568 (0.110)	0.242 (0.052)			
L21E06M *	0.936 (0.081)	-2.465 (0.221)	0.304 (0.097)			
L21E07C *	0.463 (0.018)	-0.810 (0.042)		-0.018 (0.071)	0.018 (0.055)	
L21E08M *	1.023 (0.064)	-0.749 (0.081)	0.201 (0.036)			
L21E09M *	0.620 (0.085)	0.464 (0.145)	0.340 (0.041)			
L21E10C *	0.859 (0.047)	-2.716 (0.101)				
L21E11M *	1.205 (0.074)	-0.817 (0.071)	0.226 (0.035)			
L21E12C *	1.116 (0.057)	-2.136 (0.066)				
L21E13C *	0.528 (0.027)	-0.275 (0.046)				
L21E14C *	0.493 (0.020)	0.186 (0.032)		0.852 (0.053)	-0.852 (0.053)	
L21E15C *	0.795 (0.040)	-2.219 (0.087)				
L21E16C *	0.706 (0.032)	-0.959 (0.051)				
L21E17M *	1.047 (0.065)	-0.467 (0.069)	0.171 (0.032)			
R21K01C *	0.422 (0.013)	-0.891 (0.033)		0.186 (0.052)	-0.186 (0.039)	
R21K02C *	0.807 (0.023)	-0.559 (0.025)				
R21K03M *	1.004 (0.047)	0.081 (0.039)	0.184 (0.018)			
R21K04M *	1.062 (0.096)	0.979 (0.045)	0.391 (0.014)			
R21K05C *	0.969 (0.026)	0.137 (0.017)				
R21K06M *	1.489 (0.067)	0.052 (0.029)	0.281 (0.015)			
R21K07C *	0.682 (0.016)	0.143 (0.015)		0.119 (0.027)	-0.119 (0.027)	
R21K08M *	0.994 (0.052)	0.354 (0.037)	0.197 (0.017)			
R21K09M *	1.203 (0.056)	-0.010 (0.037)	0.246 (0.018)			
R21K10C *	0.785 (0.017)	0.778 (0.015)		-0.397 (0.027)	0.397 (0.031)	
R21K11M *	1.070 (0.056)	0.260 (0.039)	0.240 (0.017)			
R21K12C *	0.576 (0.013)	-0.110 (0.014)		0.446 (0.040)	-0.084 (0.037)	-0.362 (0.034)
Items not Shared with PIRLS 2016 (Estimated Item Parameters):						
L21L01M	0.838 (0.126)	-1.433 (0.168)	0.195 (0.054)			
L21L02M	0.647 (0.138)	-0.588 (0.206)	0.217 (0.056)			
L21L03C	0.516 (0.058)	-1.925 (0.132)				
L21L04C	0.656 (0.066)	-1.881 (0.108)				

* Items with fixed item parameters estimated in PIRLS 2016 concurrent item calibration.

Item Parameters from PIRLS Literacy 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
L21L05C	0.534 (0.043)	-1.394 (0.075)		0.019 (0.143)	-0.019 (0.132)	
L21L06C	0.806 (0.078)	-2.118 (0.101)				
L21L07M	1.031 (0.146)	-1.384 (0.132)	0.182 (0.047)			
L21L08M	0.880 (0.165)	-0.776 (0.160)	0.248 (0.050)			
L21L09C	0.400 (0.054)	-1.017 (0.144)				
L21L10M	1.221 (0.186)	-1.194 (0.119)	0.221 (0.045)			
L21L11C	0.489 (0.064)	-0.125 (0.157)				
L21L12M	1.101 (0.169)	-1.518 (0.149)	0.245 (0.055)			
L21L13C	0.586 (0.038)	-1.530 (0.070)		-1.098 (0.178)	1.098 (0.167)	
L21L14M	1.666 (0.249)	-1.382 (0.096)	0.212 (0.042)			
L21L15C	0.440 (0.044)	-0.886 (0.091)		0.371 (0.159)	-0.371 (0.165)	
L21M01M	0.566 (0.089)	-3.339 (0.368)	0.241 (0.093)			
L21M02M	1.340 (0.228)	-0.245 (0.089)	0.155 (0.027)			
L21M03C	0.403 (0.054)	-3.054 (0.241)				
L21M04C	0.733 (0.070)	-1.325 (0.087)				
L21M05M	1.097 (0.164)	-2.371 (0.169)	0.234 (0.065)			
L21M06C	0.853 (0.079)	-1.889 (0.087)				
L21M07M	1.039 (0.185)	-0.605 (0.126)	0.226 (0.040)			
L21M08C	0.991 (0.094)	-2.201 (0.087)				
L21M09M	0.793 (0.118)	-1.455 (0.168)	0.178 (0.053)			
L21M10C	0.423 (0.035)	-2.004 (0.100)		-0.289 (0.193)	0.289 (0.166)	
L21M11M	1.347 (0.270)	-0.177 (0.104)	0.242 (0.032)			
L21M12C	0.677 (0.067)	-1.633 (0.098)				
L21M13M	1.109 (0.158)	-2.173 (0.150)	0.204 (0.058)			
L21M14C	0.586 (0.047)	-0.850 (0.058)		0.667 (0.133)	0.012 (0.140)	-0.679 (0.150)
L21M15M	1.710 (0.335)	-0.455 (0.095)	0.315 (0.034)			
L21M16C	0.833 (0.079)	-1.393 (0.083)				
L21M17C	1.030 (0.104)	-0.372 (0.077)				
L21M18M	1.162 (0.202)	-0.472 (0.108)	0.180 (0.035)			
L11O01M	0.905 (0.145)	-2.254 (0.226)	0.299 (0.075)			
L11O02C	0.509 (0.058)	-0.799 (0.119)				
L11O03M	1.155 (0.181)	-2.672 (0.180)	0.230 (0.069)			
L11O04M	1.168 (0.171)	-1.652 (0.139)	0.248 (0.052)			
L11O05C	0.896 (0.081)	-1.836 (0.084)				
L11O06C	0.665 (0.064)	-1.491 (0.097)				

Item Parameters from PIRLS Literacy 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
L11O07M	0.628 (0.107)	-1.644 (0.261)	0.228 (0.070)			
L11O08M	0.607 (0.104)	-1.735 (0.277)	0.231 (0.073)			
L11O09C	1.230 (0.110)	-1.684 (0.066)				
L11O10C	1.015 (0.107)	-3.050 (0.117)				
L11O11C	0.720 (0.069)	-1.883 (0.100)				
L11O12M	0.547 (0.087)	-2.109 (0.293)	0.196 (0.073)			
L11O13M	0.991 (0.169)	-0.667 (0.127)	0.200 (0.040)			
L11O14C	0.864 (0.080)	-1.884 (0.089)				
L11O15C	0.678 (0.065)	0.092 (0.086)		0.350 (0.105)	-0.350 (0.148)	
L11O16M	1.216 (0.207)	-0.420 (0.099)	0.176 (0.032)			
L11U01C	0.651 (0.065)	-2.359 (0.125)				
L11U02M	0.978 (0.155)	-2.199 (0.205)	0.291 (0.070)			
L11U03M	1.131 (0.163)	-1.283 (0.121)	0.192 (0.043)			
L11U04C	0.681 (0.066)	-1.759 (0.102)				
L11U05M	1.186 (0.159)	-1.713 (0.121)	0.177 (0.045)			
L11U06M	1.194 (0.167)	-1.405 (0.116)	0.186 (0.043)			
L11U07M	1.092 (0.207)	-0.267 (0.112)	0.198 (0.034)			
L11U08M	1.123 (0.190)	-0.732 (0.118)	0.223 (0.039)			
L11U09M	1.071 (0.176)	-1.305 (0.152)	0.278 (0.051)			
L11U10C	0.534 (0.064)	-0.304 (0.134)				
L11U11C	0.649 (0.052)	-1.701 (0.072)		0.220 (0.130)	-0.220 (0.111)	
L11U12C	0.786 (0.074)	-1.447 (0.087)				
L11U13M	1.386 (0.217)	-1.523 (0.125)	0.268 (0.049)			
L11U14C	0.594 (0.074)	-0.008 (0.141)				
L11A01M	0.958 (0.153)	-2.792 (0.234)	0.259 (0.083)			
L11A02C	0.614 (0.063)	-1.861 (0.110)				
L11A03M	0.759 (0.109)	-1.942 (0.195)	0.182 (0.061)			
L11A04C	0.792 (0.074)	-1.498 (0.084)				
L11A05M	1.228 (0.228)	-0.491 (0.111)	0.252 (0.036)			
L11A06C	0.972 (0.097)	-2.614 (0.103)				
L11A07C	0.811 (0.062)	-2.308 (0.069)		-0.050 (0.130)	0.050 (0.103)	
L11A08M	0.995 (0.140)	-1.418 (0.135)	0.181 (0.047)			
L11A09C	1.208 (0.109)	-1.792 (0.067)				
L11A10C	0.774 (0.074)	-0.964 (0.083)				
L11A11C	0.492 (0.061)	-0.325 (0.144)				

Item Parameters from PIRLS Literacy 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
L11A12CZ	0.659 (0.040)	-1.349 (0.049)		-0.418 (0.149)	0.059 (0.172)	0.358 (0.135)
L11A13CZ	0.625 (0.048)	-1.067 (0.065)		0.007 (0.120)	-0.007 (0.121)	
L11A14M	0.791 (0.126)	-1.094 (0.162)	0.163 (0.050)			
L21C01C	0.744 (0.072)	-1.738 (0.095)				
L21C02M	0.935 (0.135)	-0.997 (0.121)	0.140 (0.039)			
L21C03M	0.882 (0.192)	-0.347 (0.153)	0.255 (0.045)			
L21C04C	1.280 (0.119)	-1.917 (0.069)				
L21C05C	1.261 (0.117)	-1.896 (0.069)				
L21C06M	1.452 (0.223)	-1.213 (0.106)	0.240 (0.043)			
L21C07C	0.952 (0.100)	-2.879 (0.121)				
L21C08M	1.164 (0.203)	-0.679 (0.116)	0.236 (0.040)			
L21C09C	0.416 (0.041)	-1.536 (0.100)		0.910 (0.175)	-0.910 (0.154)	
L21C10M	1.339 (0.184)	-1.772 (0.116)	0.197 (0.049)			
L21C11C	1.095 (0.099)	-1.458 (0.069)				
L21C12C	0.675 (0.046)	-1.850 (0.068)		-0.695 (0.154)	0.695 (0.138)	
L21C13M	0.777 (0.131)	-1.145 (0.182)	0.208 (0.056)			
L21C14C	1.043 (0.099)	-2.016 (0.083)				
L21C15M	1.766 (0.284)	-0.792 (0.081)	0.223 (0.034)			
L21C16C	1.128 (0.103)	-1.246 (0.066)				
L21C17C	0.567 (0.072)	-0.105 (0.139)				
L21H01C	0.820 (0.083)	-2.706 (0.121)				
L21H02M	1.006 (0.154)	-1.771 (0.169)	0.264 (0.060)			
L21H03M	1.152 (0.200)	-0.378 (0.105)	0.180 (0.033)			
L21H04M	1.162 (0.174)	-1.756 (0.145)	0.259 (0.056)			
L21H05M	1.309 (0.196)	-1.756 (0.130)	0.260 (0.054)			
L21H06C	0.761 (0.075)	-2.195 (0.106)				
L21H07M	1.034 (0.158)	-1.118 (0.131)	0.206 (0.045)			
L21H08M	1.472 (0.253)	-0.565 (0.095)	0.241 (0.034)			
L21H09M	1.111 (0.161)	-1.391 (0.129)	0.210 (0.048)			
L21H10M	0.784 (0.140)	-0.873 (0.174)	0.216 (0.052)			
L21H11C	0.558 (0.060)	-1.605 (0.114)				
L21H12C	0.680 (0.068)	-1.299 (0.094)				
L21H13M	1.366 (0.213)	-1.162 (0.112)	0.248 (0.043)			
L21H14C	0.811 (0.077)	-1.702 (0.088)				
L21H15M	1.443 (0.297)	-0.229 (0.104)	0.279 (0.033)			

Item Parameters from PIRLS Literacy 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
L21H16M	1.115 (0.197)	-0.751 (0.130)	0.254 (0.043)			
L11P01M	0.939 (0.163)	-0.866 (0.149)	0.237 (0.046)			
L11P02M	0.938 (0.155)	-1.114 (0.156)	0.244 (0.050)			
L11P03C	0.671 (0.056)	-1.014 (0.066)		0.625 (0.106)	-0.625 (0.110)	
L11P04C	0.845 (0.083)	-2.581 (0.108)				
L11P05M	0.997 (0.146)	-1.540 (0.149)	0.220 (0.052)			
L11P06C	0.803 (0.081)	-2.732 (0.119)				
L11P07C	0.733 (0.071)	-0.856 (0.089)				
L11P08M	1.390 (0.200)	-1.184 (0.099)	0.190 (0.038)			
L11P09M	1.330 (0.205)	-1.436 (0.123)	0.265 (0.047)			
L11P10M	1.325 (0.208)	-1.383 (0.124)	0.273 (0.047)			
L11P11C	0.642 (0.044)	-1.237 (0.064)		-0.431 (0.133)	0.431 (0.131)	
L11P12C	0.565 (0.062)	-0.803 (0.112)				
L11P13C	1.128 (0.102)	-1.167 (0.067)				
L11P14C	0.606 (0.065)	-1.032 (0.105)				

Appendix 12C: ePIRLS 2016 Item Parameters from Item Calibration

Item Parameters from ePIRLS 2016 Item Calibration

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
ePIRLS 2016 Items (Estimated Item Parameters):						
E11B01M	0.583 (0.058)	-1.033 (0.241)	0.245 (0.079)			
E11B02M	0.982 (0.077)	-0.918 (0.130)	0.264 (0.063)			
E11B03C	0.681 (0.042)	1.037 (0.051)				
E11B04C	0.539 (0.035)	-0.719 (0.072)				
E11B05M	1.198 (0.092)	0.600 (0.042)	0.137 (0.022)			
E11B06C	0.683 (0.028)	0.348 (0.023)		0.077 (0.043)	-0.077 (0.043)	
E11B07M	0.788 (0.076)	0.364 (0.094)	0.177 (0.040)			
E11B08C	1.126 (0.050)	-0.447 (0.033)				
E11B09C	1.177 (0.050)	-0.226 (0.028)				
E11B10C	0.601 (0.029)	0.941 (0.034)		0.168 (0.045)	-0.168 (0.057)	
E11B11M	1.217 (0.118)	1.007 (0.042)	0.156 (0.019)			
E11B12C	1.035 (0.049)	0.668 (0.027)				
E11B13C	1.191 (0.051)	-0.041 (0.025)				
E11B14C	0.924 (0.046)	0.709 (0.030)				
E11B15C	0.558 (0.025)	0.025 (0.030)		-0.041 (0.057)	0.041 (0.051)	
E11B16C	0.288 (0.014)	-0.274 (0.047)		-0.830 (0.140)	0.375 (0.135)	0.454 (0.106)
E11B17C	0.437 (0.021)	0.436 (0.028)		0.156 (0.081)	0.379 (0.079)	-0.535 (0.078)
E11M01M	1.302 (0.125)	1.040 (0.040)	0.158 (0.018)			
E11M02C	0.805 (0.049)	-1.633 (0.095)				
E11M03C	0.616 (0.036)	-0.470 (0.054)				
E11M04C	1.111 (0.048)	-0.035 (0.026)				
E11M05M	1.493 (0.130)	-0.438 (0.089)	0.458 (0.044)			
E11M06M	0.834 (0.088)	0.213 (0.116)	0.277 (0.046)			
E11M07M	1.300 (0.111)	0.806 (0.039)	0.165 (0.020)			
E11M08C	1.027 (0.047)	-0.448 (0.036)				
E11M09C	0.598 (0.035)	-0.050 (0.044)				
E11M10M	1.349 (0.090)	-0.124 (0.057)	0.200 (0.034)			
E11M11C	0.534 (0.027)	0.779 (0.034)		0.462 (0.049)	-0.462 (0.059)	
E11M12M	1.229 (0.104)	0.195 (0.068)	0.298 (0.034)			
E11M13C	0.900 (0.049)	1.101 (0.042)				

Item Parameters from ePIRLS 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
E11M14C	0.616 (0.020)	0.530 (0.019)		-0.450 (0.059)	0.651 (0.062)	-0.200 (0.054)
E11M15C	0.967 (0.048)	-0.702 (0.046)				
E11M16C	0.753 (0.023)	0.164 (0.021)		-0.567 (0.048)	0.567 (0.046)	
E11M17C	0.432 (0.019)	0.526 (0.034)		-0.628 (0.071)	0.628 (0.074)	
E11M18C	0.902 (0.045)	0.470 (0.029)				
E11M19M	1.376 (0.116)	0.658 (0.043)	0.190 (0.024)			
E11M20C	0.818 (0.050)	1.169 (0.049)				
E11R01M	0.829 (0.068)	-0.314 (0.115)	0.203 (0.050)			
E11R02C	0.588 (0.037)	-1.276 (0.094)				
E11R03C	0.493 (0.024)	0.179 (0.032)		0.396 (0.058)	-0.396 (0.055)	
E11R04M	1.662 (0.137)	0.861 (0.031)	0.187 (0.017)			
E11R05C	0.666 (0.037)	-0.313 (0.046)				
E11R06C	0.718 (0.042)	-0.861 (0.067)				
E11R07C	0.994 (0.033)	0.400 (0.017)		-0.065 (0.032)	0.065 (0.032)	
E11R08C	0.757 (0.041)	0.647 (0.035)				
E11R09C	0.680 (0.037)	0.111 (0.037)				
E11R10M	1.084 (0.097)	0.085 (0.087)	0.316 (0.040)			
E11R11C	0.682 (0.029)	0.437 (0.024)		0.214 (0.042)	-0.214 (0.043)	
E11R12M	1.749 (0.122)	0.048 (0.045)	0.274 (0.029)			
E11R13M	0.756 (0.087)	0.130 (0.148)	0.306 (0.053)			
E11R14C	0.829 (0.041)	-0.094 (0.035)				
E11R15C	1.247 (0.053)	0.064 (0.024)				
E11R16C	0.787 (0.042)	0.236 (0.033)				
E11T01M	0.758 (0.099)	0.451 (0.134)	0.315 (0.048)			
E11T02C	1.158 (0.060)	-1.152 (0.053)				
E11T03M	0.865 (0.080)	0.387 (0.081)	0.176 (0.036)			
E11T04M	1.404 (0.093)	-0.439 (0.065)	0.217 (0.040)			
E11T05C	0.863 (0.032)	-0.356 (0.025)		0.085 (0.044)	-0.085 (0.034)	
E11T06C	1.013 (0.045)	0.182 (0.026)				
E11T07M	0.977 (0.131)	1.320 (0.065)	0.175 (0.022)			
E11T08C	0.734 (0.038)	-0.060 (0.037)				
E11T09C	0.934 (0.050)	1.070 (0.040)				
E11T10C	1.102 (0.048)	0.269 (0.024)				
E11T11M	0.946 (0.061)	-0.456 (0.078)	0.120 (0.038)			
E11T12M	0.737 (0.074)	-0.451 (0.180)	0.296 (0.067)			

Item Parameters from ePIRLS 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
E11T13M	1.468 (0.127)	0.433 (0.051)	0.323 (0.027)			
E11T14C	0.805 (0.043)	0.678 (0.034)				
E11T15M	1.955 (0.121)	0.282 (0.030)	0.166 (0.020)			
E11T16M	1.270 (0.095)	-0.502 (0.086)	0.275 (0.049)			
E11T17C	1.082 (0.048)	0.169 (0.025)				
E11T18C	0.975 (0.049)	0.732 (0.030)				
E11Z01M	0.969 (0.098)	-0.248 (0.136)	0.411 (0.052)			
E11Z02C	0.511 (0.023)	0.260 (0.029)		-0.043 (0.057)	0.043 (0.056)	
E11Z03M	0.816 (0.079)	0.220 (0.102)	0.211 (0.043)			
E11Z04C	1.068 (0.053)	-0.954 (0.049)				
E11Z05M	0.627 (0.068)	-0.082 (0.174)	0.225 (0.060)			
E11Z06C	1.356 (0.055)	0.229 (0.020)				
E11Z07M	1.403 (0.099)	0.081 (0.052)	0.228 (0.030)			
E11Z08M	1.133 (0.100)	0.261 (0.072)	0.284 (0.034)			
E11Z09C	0.620 (0.041)	1.212 (0.064)				
E11Z10M	1.202 (0.090)	0.344 (0.051)	0.175 (0.027)			
E11Z11M	1.238 (0.091)	-0.650 (0.091)	0.272 (0.051)			
E11Z12C	0.784 (0.039)	-0.083 (0.035)				
E11Z13M	1.429 (0.097)	-0.369 (0.064)	0.232 (0.039)			
E11Z14C	0.884 (0.034)	-0.498 (0.028)		0.134 (0.047)	-0.134 (0.033)	
E11Z15M	1.291 (0.101)	0.193 (0.059)	0.255 (0.032)			
E11Z16C	0.880 (0.033)	0.146 (0.020)		0.151 (0.036)	-0.151 (0.033)	
E11Z17C	0.710 (0.042)	0.838 (0.042)				
E11Z18M	1.180 (0.096)	0.145 (0.069)	0.245 (0.036)			
E11Z19C	1.035 (0.040)	0.599 (0.018)		0.338 (0.028)	-0.338 (0.032)	
E11Z20C	0.963 (0.037)	0.072 (0.020)		0.214 (0.036)	-0.214 (0.031)	

CHAPTER 13

Using Scale Anchoring to Interpret the PIRLS and ePIRLS 2016 Achievement Scales

Ina V.S. Mullis
Caroline O. Prendergast

Introduction

As described in Chapter 12: Scaling the PIRLS 2016 Achievement Data, the PIRLS 2016 achievement results are summarized using item response theory (IRT) scaling and reported on achievement scales, with most achievement scores ranging from 300 to 700. Countries' average scores provide users of the data with information about how achievement compares among countries and whether scores are improving or declining over time.

To provide as much information as possible for policy and curriculum reform, however, it is important to understand the reading competencies associated with different locations within the range of scores on the achievement scales. For example, in terms of levels of student understanding, what does it mean for a country to have average achievement of 513 or 426, and how are these scores different?

The PIRLS 2016 International Benchmarks provide information about what students know and can do at different points along the PIRLS achievement scale. More specifically, PIRLS has identified four points along the PIRLS achievement scale to use as international benchmarks of achievement—Advanced International Benchmark (625), High International Benchmark (550), Intermediate International Benchmark (475), and Low International Benchmark (400).

The TIMSS & PIRLS International Study Center worked with the expert international committee, the Reading Development Group (RDG), to conduct two scale anchoring analyses to describe student competencies at each of the benchmarks for PIRLS 2016 and ePIRLS 2016, respectively.

This chapter details the scale anchoring procedures that were followed to describe student performance at the international benchmarks for PIRLS and ePIRLS 2016. In brief, scale anchoring involved identifying items that students scoring at an international benchmark answered correctly, and then having experts examine the content of each item to determine the reading comprehension skills and strategies demonstrated by students who responded correctly to the item. The experts then summarized the detailed list of item competencies in a brief description of achievement at each international benchmark. Thus, the scale anchoring procedure yielded a content-referenced interpretation of the achievement results that can be considered in light of the PIRLS 2016 frameworks for assessing reading. The first scale anchoring analysis was conducted for the PIRLS and PIRLS Literacy items to benefit from the whole range of PIRLS items (see Chapter 1). The second analysis was for the ePIRLS items, which measure online informational reading.

PIRLS and PIRLS Literacy

PIRLS Literacy was introduced in 2016 to provide an extension of PIRLS for countries where most children in the fourth grade are still developing fundamental reading skills (see the *PIRLS 2016 Assessment Framework*). The PIRLS Literacy passages are shorter with less complex syntax than the PIRLS passages, and the questions are presented side by side to support the location of information. So that PIRLS Literacy could be reported on the PIRLS achievement scale, the two versions (PIRLS and PIRLS Literacy) have four passages in common.

In order to take full advantage of the information offered by PIRLS Literacy, items from both versions of the assessment were included in the scale anchoring process. PIRLS items and items from the four passages common to PIRLS and PIRLS Literacy were considered at all four benchmarks using data from the students who participated in PIRLS. Passages and items that appeared only in PIRLS Literacy were considered only for the Low and Intermediate benchmarks. Because the PIRLS Literacy items were developed to improve measurement at the lower end of the PIRLS scale, their inclusion in the scale anchoring process provided necessary information about the abilities of students reaching the Low and Intermediate benchmarks.

ePIRLS

PIRLS was extended during the 2016 cycle to include ePIRLS, which uses a simulated Internet environment to measure online informational reading. The five ePIRLS tasks asked students to navigate through interconnected webpages containing both textual and visual information to complete school-like assignments about science and social studies topics. Because ePIRLS assesses reading comprehension skills that are specific to informational reading in an Internet environment, the scale anchoring process for ePIRLS was conducted separately from the scale anchoring process for the informational items included in PIRLS. This resulted in benchmark descriptions that were specific to the complex demands of online reading.

Classifying the Items

As the first step, students scoring at the four benchmarks were identified for analysis. For PIRLS 2016, students scoring within 5 scale-score points of each benchmark (i.e., the benchmark point plus or minus 5) were identified for the benchmark analysis. This 10-point range provided an adequate sample of students scoring at the benchmark, and yet was small enough so that performance at one international benchmark was still distinguishable from the next. For passages and items that were included in both PIRLS and PIRLS Literacy, scale anchoring was conducted using data from the students who participated in PIRLS. For passages and items included only in PIRLS Literacy, the range was expanded to students scoring within 10 scale-score points of each benchmark in order to obtain a sufficiently large sample. The 10-point range (the benchmark point plus or minus 5) was used for ePIRLS, including the students that had participated in both PIRLS and ePIRLS. The score ranges around each international benchmark and the number of students scoring in each range are shown in Exhibit 13.1.

Exhibit 13.1: Range Around Each International Benchmark and Number of Students Within Each Range

	Low (400)	Intermediate (475)	High (550)	Advanced (625)
PIRLS (including passages also in PIRLS Literacy)				
Range of Scale Scores	395–405	470–480	545–555	620–630
Number of Students	3,556	8,198	12,905	6,882
PIRLS Literacy (only)				
Range of Scale Scores	390–410	465–485	540–560	615–635
Number of Students	2,305	1,765	925	232
ePIRLS				
Range of Scale Scores	395–405	470–480	545–555	620–630
Number of Students	1,000	2,229	3,711	2,189

The second step involved computing the percentage of those students scoring in the range around each international benchmark that answered each item correctly. To compute these percentages, students in each country were weighted proportionally to the size of the student population in the country. For multiple-choice items and constructed response items worth 1 point, it was a straightforward matter of computing the percentage of students at each benchmark who answered each item correctly. For constructed response items and compound multiple-choice items scored for partial and full credit, percentages were computed for students receiving partial credit (1 point or 2 points) as well as for students receiving full credit (2 points or 3 points).

Third, the criteria described below were applied to identify the items that anchored at each benchmark. An important feature of the scale anchoring method is that it yields descriptions of the performance demonstrated by students reaching each of the international benchmarks on the scale, and that the descriptions reflect demonstrably different accomplishments by students reaching each successively higher benchmark. Because the process entails the delineation of sets of items that students at each international benchmark are likely to answer correctly and that discriminate between one benchmark and the next, the criteria for identifying the items that anchor considers performance at more than one benchmark.

For multiple-choice items, 65 percent was used as the criterion for anchoring at each benchmark being analyzed, since students would be likely (about two thirds of the time) to answer the item correctly. A somewhat less strict criterion was used for the constructed response items, because students had much less scope for guessing. For constructed response items, the criterion of 50 percent was used for the benchmark without any discrimination criterion for the next lower benchmark. In addition, a criterion of less than 50 percent was used for the next lower benchmark, because with this response probability, students were more likely to have answered the item incorrectly than correctly.

Using a multiple-choice item as an example, the criteria for each benchmark are outlined below:

- A multiple-choice item anchored at the Low International Benchmark (400) if at least 65 percent of students scoring in the range answered the item correctly. Because this was the lowest benchmark described, there were no further criteria.
- A multiple-choice item anchored at the Intermediate International Benchmark (475) if at least 65 percent of students scoring in the range answered the item correctly, and less than 50 percent of the students at the Low International Benchmark answered the item correctly.
- A multiple-choice item anchored at the High International Benchmark (550) if at least 65 percent of students scoring in the range answered the item correctly, and less than 50 percent of students at the Intermediate International Benchmark answered the item correctly.
- A multiple-choice item anchored at the Advanced International Benchmark (625) if at least 65 percent of students scoring in the range answered the item correctly, and less than 50 percent of students at the High International Benchmark answered the item correctly.

To include all of the multiple-choice items in the anchoring process and provide information about comprehension processes that might not otherwise have had many anchor items, the concept of items that "almost anchored" was introduced. These were items that met slightly less stringent criteria for being answered correctly. The criteria to identify multiple-choice items that "almost anchored" were that 60 to 65 percent of students scoring in the range answered the item correctly and less than 50 percent of students at the next lowest benchmark answered the item correctly. To be completely inclusive for all items, items that met only the criterion that 60 to 65 percent of the students answered correctly (regardless of the performance of students at the next lower point) were also identified. The categories of items were mutually exclusive, and ensured that all of the items were available to inform the descriptions of student achievement at the anchor levels. A multiple-choice item was considered to be "too difficult" to anchor if less than 60 percent of students at the advanced benchmark answered the item correctly. A constructed response item was considered to be "too difficult" to anchor if less than 50 percent of students at the advanced benchmark answered the item correctly.

Exhibit 13.2 presents the number of PIRLS 2016 items that anchored at each international benchmark.

Exhibit 13.2: Number of Items Anchoring and Almost Anchoring at Each International Benchmark*

	Low (400)	Intermediate (475)	High (550)	Advanced (625)
PIRLS Literary	62	39	48	17
PIRLS Informational	54	35	48	29
ePIRLS	15	24	36	22

* Item counts for PIRLS Literary and PIRLS Informational include items that appeared only in PIRLS Literacy that anchored at the Low and Intermediate International Benchmarks.

Preparing the Scale Anchoring Documentation

The scale anchoring for PIRLS and ePIRLS 2016 was conducted in the spring of 2017 at a four-day meeting in Lübeck, Germany. To prepare documentation for use by the RDG, staff at the TIMSS & PIRLS International Study prepared short descriptions of the student competencies demonstrated by a correct (or partially correct) response to each item. The descriptions were updated for trend items from previous assessment cycles, and new descriptions were drafted for the items assessed for the first time in 2016. Complete documentation provided for each item included the description, framework classification, answer key or scoring guide, secure status, scale anchoring data, and international mean. An example scale anchoring page for an item at the Intermediate Benchmark is presented in Appendix 13A.

The items, scoring guides, and documentation were grouped by reading purpose (for the PIRLS scale anchoring analysis) and then by international benchmark. The final categorization was by the anchoring criteria the items met—items that anchored, followed by items that almost anchored, then by items that met only the 60 to 65 percent criteria.

At the scale anchoring meeting, the expert committee 1) worked through each item to finalize the description of the student competencies demonstrated by a correct (or partially correct) response, 2) summarized the proficiency demonstrated by students reaching each international benchmark for publication in reports, and 3) selected example items that supported and illustrated the benchmark descriptions to publish together with the descriptions.

Following the scale anchoring meeting, the descriptions and example items published in the PIRLS 2016 reports were reviewed by National Research Coordinators at their 8th meeting in Riga, Latvia. Appendix 13B contains the scale anchoring descriptions for the PIRLS literary items, Appendix 13C contains the scale anchoring descriptions for the PIRLS informational items, and Appendix 13D contains the scale anchoring descriptions for the ePIRLS items. Scale anchoring considered partial credit and full credit responses separately. Because of this, a partial credit item can anchor more than once, typically at a higher benchmark for full credit, and a lower benchmark for partial credit. If they both anchored at the same level, the full credit results were used for the analysis.

Appendix 13A: Sample Scale Anchoring Page for Item at Intermediate Benchmark

ID: R41H11M	Why does Macy make white wings on a pole	Block_Seq: H_11

11. Why does Macy make white wings on a pole?

Ⓐ to look like hen's feathers

Ⓑ to make a decision

Ⓒ to look like an owl

Ⓓ to impress Sam

Purpose

Literary Experience

Process

Make Straightforward Inferences

Points

1

Format

MC

Key

C

Secure Status

Restricted Use

Proposed 2016 Scale Anchoring Description

Infer and recognize the reason for a character's action

Weighted Percent Correct at International Benchmark Levels

Low (400)	Intermediate (475)	High (550)	Advanced (625)	Intermediate
42	75	92	100	International % Correct: 79

2016 Scale Anchoring 123

Appendix 13B: PIRLS 2016 Literary Item Descriptions Developed During the PIRLS 2016 Benchmarking

Items at Low International Benchmark (400)		
P/PL	F_10	Recognize and reproduce a character's feeling that is clearly suggested at a specified point in the story
P/PL	B_01	Retrieve and reproduce explicitly stated information about the central character
P/PL	B_02	Retrieve the explicitly stated reason for a character's action
P/PL	B_03	Retrieve the explicitly stated reason for a character's action
P/PL	B_04	Make a straightforward inference about a reason for a character's action
P/PL	B_06	Make a straightforward inference about the reason for an event
P/PL	B_07	Retrieve and reproduce an explicitly stated detail about the reason for an event
P/PL	B_08	Make a straightforward inference about the purpose of a character's action
P/PL	B_12	Make an inference to recognize the purpose of a character's action
P/PL	B_13	Interpret story events to determine the cause of one of a character's stated feelings
P	M_09	Reproduce a straightforward reason for an action
P	H_01	Locate an explicitly stated character action from the beginning of the text
PL	M_01	Locate explicitly stated information at the beginning of the text
PL	M_03	Locate and reproduce an explicitly stated action of a character
PL	M_04	Locate and reproduce an explicitly stated detail
PL	M_05	Recognize and retrieve an explicitly stated detail
PL	M_06	Recognize and reproduce explicitly stated information
PL	M_08	Locate and reproduce explicitly stated information
PL	M_09	Make an inference about the reason for an event
PL	M_10	Locate and reproduce 2 explicitly stated feelings of a character
PL	M_12	Locate and reproduce an explicitly stated detail
PL	M_13	Locate and make a straightforward inference about a character's action
PL	M_16	Locate and reproduce an explicitly stated action
PL	O_01	Locate explicitly stated information at the beginning of the text
PL	O_03	Locate and recognize explicitly stated information
PL	O_04	Retrieve an explicitly stated character trait

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

PL	O_05	Locate and reproduce a character's idea
PL	O_06	Make a straightforward inference about a character's words
PL	O_07	Make a straightforward inference about the purpose for a character's action
PL	O_08	Make a straightforward inference about a character's feeling
PL	O_09	Locate and reproduce explicitly stated information
PL	O_10	Locate and reproduce explicitly stated information
PL	O_11	Make a straightforward inference about a character's reaction
PL	O_12	Make a straightforward inference about a character's reaction
PL	O_14	Locate and reproduce explicitly stated information
PL	U_01	Locate and reproduce explicitly stated information at the beginning of the text
PL	U_02	Locate and recognize an explicitly stated action
PL	U_03	Make a straightforward inference about the reason for a situation
PL	U_04	Locate and reproduce the reason for a situation
PL	U_05	Locate and recognize an explicitly stated reason for a character's action
PL	U_06	Locate and recognize the explicitly stated reason for a situation
PL	U_09	Make a straightforward inference about the reason for a character's action
PL	U_11	Locate and reproduce 1 (of 2) pieces of explicitly stated information
PL	U_11	Locate and reproduce 2 pieces of explicitly stated information
PL	U_12	Determine the sequence of events of the whole story
PL	U_13	Evaluate the whole story and recognize a central idea
PL	L_01	Make a straightforward inference about a detail from the beginning of the story
PL	L_03	Locate and reproduce an explicitly stated reason for a character's words
PL	L_04	Locate and reproduce the reason for a character's words
PL	L_05	Make a straightforward inference and reproduce 1 (of 2) of a character's actions
PL	L_06	Retrieve and reproduce explicitly stated information
PL	L_07	Retrieve and recognize explicitly stated explanation of a character's action
PL	L_09	Integrate ideas to show understanding of how a character develops
PL	L_12	Retrieve an explicitly stated reason for a character's action
PL	L_13	Locate and reproduce 1 (of 2) explicitly stated detail

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

PL	L_14	Locate and recognize an explicitly stated idea
PL	L_15	Show understanding of a character's trait by 1 (of 2) example of a character's actions
P	O_01	Retrieve and recognize a character's expectations about a future event
P/PL	B_05	Locate and recognize an embedded detail
P/PL	B_11	Retrieve and recognize an explicitly stated cause of a character's action
Items at Intermediate International Benchmark (475)		
P/PL	F_01	Identify the narrator (in a first person story) from a range of clues in the text and confirmed by the pictures
P/PL	F_04	Retrieve and recognize explicitly stated information
P/PL	F_06	Make an inference to explain a character's reaction to an event
P/PL	F_09	Reproduce 1 (of 2) explicitly stated character action
P	O_02	Recognize and reproduce explicitly stated information
P	O_07	Infer 2 physical characteristics from a description
P	Y_09	Locate a central event and make a straightforward inference to provide 1 (of 2) character action
P	Y_13	Interpret and integrate story events and character actions to describe or illustrate a character trait
P/PL	B_09	Locate and reproduce an explicit action from a sequence
P/PL	B_16	Locate and reproduce 1 (of 2) action to give a reason for a character's change in thinking
P/PL	B_17	Integrate ideas across the text to provide a character description or action
P	M_02	Locate and reproduce an explicit detail embedded in the introductory paragraph
P	M_08	Locate and retrieve an explicit action from a sequence
P	M_11	Locate and retrieve an explicitly stated feeling
P	M_13	Make a straightforward inference about a character's reaction to a situation
P	M_17	Locate and reproduce a straightforward story event as the cause of 1 (of 3) feeling
P	H_02	Recognize how an author demonstrates a character's traits
P	H_06	Locate and reproduce 1 (of 2) action that leads to a specified result
P	H_11	Infer and recognize the reason for a character's action
PL	M_07	Make a straightforward inference about a character's reaction
PL	M_14	Integrate evidence to make a causal inference
PL	M_18	Evaluate the whole story and recognize the central idea
PL	O_02	Make a straightforward inference about the reason for a character's reaction

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

PL	O_13	Integrate evidence and recognize the reason for a character's reaction
PL	O_15	Integrate evidence and reproduce either a character's reaction or an explanation for this reaction
PL	U_08	Make a straightforward inference about the reason for a character's action
PL	L_05	Make a straightforward inference and reproduce 2 of a character's actions
PL	L_13	Locate and reproduce 2 explicitly stated details
P/PL	B_10	Recognize the meaning of a simile
P/PL	B_14	Locate and integrate evidence to recognize a character's reaction
P	M_07	Recognize the reason for characters' actions
P	H_12	Integrate evidence to recognize the reason for a character's action
P/PL	F_02	Retrieve, combine, and visualize concrete descriptive information and identify matching picture
P/PL	F_03	Infer reason for an opinion from a dialogue
P	Y_07	Retrieve and recognize a character's plan of action
P	M_01	Recognize explicit central information from the introductory paragraph
PL	L_08	Locate and recognize an explicitly stated piece of information
PL	L_10	Locate and recognize an explicitly stated reason for a character's opinion
Items at High International Benchmark (550)		
P/PL	F_07	Give a simple interpretation of a character's feelings about the setting
P/PL	F_08	Infer the significance of a character's action from subsequent events
P/PL	F_09	Reproduce 2 explicitly stated character actions from different parts of the text
P/PL	F_12	Interpret the narrator's feelings at either the beginning or the end of the story
P	O_05	Locate and reproduce 1 (of 2) explicitly stated physical attribute of a character embedded in a longer description
P	O_08	Integrate ideas across text to interpret the reasons for a character's feelings
P	O_09	Interpret the reason for a character's reaction
P	O_10	Interpret and integrate a character's actions, including at least 1 character trait and 1 supporting action
P	O_13	Interpret and integrate story events to do 1 of the following: determine the reason for a character's inability to perform an action, identify another character's action that changes this, and show understanding of how this action changes another character's feelings
P	Y_01	Infer from complex imagery how a character's appearance suggests her name
P	Y_04	Locate and retrieve an embedded detail
P	Y_06	Locate and retrieve information from a dialogue within a description of a character's actions

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

P	Y_09	Locate a central event and make a straightforward inference to provide 2 character actions
P	Y_10	Interpret the motivation for a character's words by providing an example from the story
P	Y_11	Locate a relevant part of the text and recognize the meaning of a metaphor
P	Y_12	Integrate story events to support a chosen character description with evidence
P	Y_14	State a title preference based on evaluating story events and characters' actions and explain the choice in terms of the significance or central role of the character
P/PL	B_13	Interpret story events to determine the cause of two contradictory stated feelings
P/PL	B_15	Interpret the reason for a character's words
P/PL	B_17	Integrate ideas across the text to provide a character description and supporting action
P	M_03	Recognize the meaning of a metaphor central to the story
P	M_06	Show understanding of a character by examining a series of the character's actions
P	M_12	Interpret a character's hidden motivation in the context of the whole story
P	M_14	Integrate evidence from across the text to interpret the reason for a situation
P	M_15	Evaluate a character's actions across the text to interpret his underlying values
P	M_16	Show understanding of the story plot by interpreting a character's hidden intention
P	M_17	Locate and reproduce events from different parts of the story as the cause of 2 (of 3) feelings
P	H_03	Infer the reason for a character's feelings
P	H_06	Locate and reproduce 2 actions that lead to a specific result
P	H_10	Locate and recognize the inspiration for a character's idea
P	H_13	Interpret ideas from across the text to identify a character trait
P	H_14	Integrate evidence from across the text to describe a central idea
P	H_15	Integrate events across the text to predict a character's future behavior
P	H_09	Make a straightforward inference about the reason for a character's words
P/PL	F_05	Make an inference to recognize the main character's feelings
P/PL	F_11	Evaluate the tone of the story and recognize that a humorous ending fits the story
P/PL	F_13	Evaluate the whole story to recognize a central idea
P	O_06	Locate and retrieve dialogue that results in a given character emotion
P	O_11	Understand the meaning of figurative language
P	O_12	Make a straightforward inference to recognize the reason for a character's action
P	Y_02	Interpret and generalize to recognize a summary of a character's attributes

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

P	Y_05	Infer a character trait from a character's action
P	Y_08	Locate a relevant point in the story and make an inference about an event
P	M_05	Retrieve, combine, and visualize a procedural sequence and recognize matching diagram
P	H_05	Make an inference to explain a character's action
Items at Advanced International Benchmark (625)		
P/PL	F_07	Integrate ideas across text to interpret the character's feelings about the setting
P/PL	F_12	Interpret the change in the narrator's feelings between the beginning and the end of the story
P	O_03	Recognize that the author's choice of words raises suspense
P	O_05	Locate and reproduce 2 explicitly stated physical attributes of a character embedded in a longer description
P	O_10	Interpret and integrate a character's actions, including at least 1 character trait and 2 supporting actions
P	O_13	Interpret and integrate story events to fully explain the implications of the central character's problem and its resolution
P	Y_03	Infer an explanation by examining description and imagery
P	Y_13	Interpret and integrate story events and character actions to describe a character with two supporting details from the text
P/PL	B_16	Locate and reproduce 2 actions to give a reason for a character's change in thinking
P	M_04	Make an inference from a specified point in the story to find evidence to support a given description of a character
P	M_10	Interpret a possible motivation for characters' advice
P	M_17	Locate and reproduce events from different parts of the story as the cause of each of 3 feelings
P	H_07	Locate, determine, and recognize the reason for a character's point of view
P	H_13	Interpret ideas from across the text to identify and support a character trait with 1 (of 2) example
P	H_16	Evaluate story events and character actions to explain why an alternative, given title would be appropriate
Items Above the Advanced International Benchmark (625)		
P	O_04	Evaluate and reproduce 2 examples of character's words that convey an emotion
P	M_09	Contrast two situations in the story to give a reason for characters' actions
P	H_04	Interpret story events to determine the implicit reason for a character's actions
P	H_08	Evaluate and determine the implicit meaning of a character's statement
P	H_13	Integrate ideas from across the text to identify a character trait and support it with 2 examples

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

Appendix 13C: PIRLS 2016 Informational Item Descriptions Developed During the PIRLS 2016 Benchmarking

Items at Low International Benchmark (400)		
P/PL	K_01	Retrieve and reproduce 1 (of 2) piece of explicitly stated information when directed to the beginning of the text
P	I_01	Locate explicitly stated information at the beginning of the text
P	L_01	Locate explicitly stated information at the beginning of the text
P/PL	E_01	Locate and reproduce explicitly stated information from the beginning of the text
P/PL	E_02	Locate and reproduce explicitly stated information from the beginning of the text
P/PL	E_05	Retrieve and recognize an explicitly stated reason for an action
P/PL	E_06	Retrieve and recognize an explicitly stated detail
P/PL	E_07	Locate and reproduce 1 (of 2) explicitly stated detail
P/PL	E_10	Retrieve and reproduce an explicitly stated detail
P/PL	E_12	Retrieve and reproduce an explicitly stated detail
P/PL	E_15	Retrieve and reproduce an explicitly stated detail
P	W_01	Retrieve and reproduce 1 (of 2) piece of information from the beginning of the text
PL	H_01	Retrieve and reproduce an explicitly stated detail from the beginning of the text
PL	H_02	Retrieve and recognize an explicitly stated detail from the beginning of the text
PL	H_04	Make a straightforward inference about the relationship between two actions
PL	H_05	Retrieve and recognize an explicitly stated detail
PL	H_06	Retrieve and reproduce an explicitly stated detail
PL	H_11	Retrieve and reproduce an explicitly stated detail
PL	H_12	Make a straightforward inference about an expectation
PL	H_14	Retrieve and reproduce an explicitly stated detail
PL	P_03	Identify and reproduce essential information from the beginning of the text
PL	P_04	Retrieve and reproduce an explicitly stated detail
PL	P_05	Locate and infer an explanation from explicitly stated information
PL	P_06	Retrieve and reproduce explicitly stated information
PL	P_09	Locate and recognize an explicitly stated detail
PL	P_10	Locate and integrate information to recognize the significance of an action

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

PL	P_11	Locate and reproduce 1 (of 2) explicitly stated piece of information
PL	A_01	Retrieve and recognize an explicitly stated detail from the beginning of the text
PL	A_02	Retrieve and reproduce an explicitly stated detail
PL	A_03	Locate and recognize an explicitly stated detail
PL	A_04	Make a straightforward inference to reproduce a detail
PL	A_06	Retrieve and reproduce an explicitly stated detail
PL	A_07	Retrieve and reproduce 2 explicitly stated details
PL	A_08	Make a straightforward inference about a description
PL	A_09	Make a straightforward inference about a description
PL	A_10	Make a straightforward inference about an action
PL	A_12	Integrate details from across the text to complete a table (2 of 3)
PL	A_13	Locate and decide the accuracy of 3 (of 4) details from a description
PL	C_01	Retrieve and reproduce a detail from the beginning of the text
PL	C_04	Retrieve and reproduce a detail from a chart
PL	C_05	Retrieve and reproduce a detail from a chart
PL	C_07	Retrieve and reproduce an explicitly stated detail
PL	C_09	Interpret information to provide a partial explanation
PL	C_10	Locate and recognize an explicitly stated detail from a text box
PL	C_11	Retrieve and reproduce an explicitly stated detail
PL	C_12	Retrieve and reproduce 2 explicitly stated details
PL	C_14	Retrieve and reproduce an explicitly stated detail
PL	H_07	Make a straightforward inference about an explanation
PL	H_09	Retrieve and recognize an explicitly stated detail
PL	P_08	Retrieve and recognize an explicitly stated detail
PL	C_06	Make a straightforward inference about an action
Items at Intermediate International Benchmark (475)		
P/PL	K_02	Locate and reproduce 3 pieces of explicitly stated information
P/PL	K_12	Interpret and integrate information from across different sections to partially complete a table (3/6 entries)
P	I_08	Retrieve and recognize an explicitly stated definition

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

P	L_04	Locate and reproduce 1 or 2 (of 4) pieces of explicitly stated information
P	L_09	Retrieve and recognize an explicitly stated detail embedded in continuous text
P/PL	E_08	Interpret and recognize the significance of an invention
P/PL	E_11	Make a straightforward inference to recognize an explanation
P/PL	E_14	Make a straightforward inference to provide 1 (of 2) comparison
P/PL	E_16	Integrate information across text to order a set of events
P	W_04	Locate and reproduce 2 pieces of explicitly stated information from a text box
P	T_02	Locate and reproduce 1 (of 2) action that is part of a sequence of events
P	T_06	Make a straightforward inference about the cause of a situation
PL	H_16	Interpret the whole text to recognize the reason for its title
PL	P_07	Make a straightforward inference to provide an explanation
PL	P_11	Locate and reproduce 2 explicitly stated pieces of information
PL	P_12	Retrieve and reproduce an explicitly stated detail
PL	P_13	Retrieve and reproduce an explicitly stated detail
PL	P_14	Integrate information to order a set of events
PL	A_12	Integrate details from across the text to complete a table (3 of 3)
PL	A_13	Locate and decide the accuracy of 4 (of 4) details from a description
PL	C_02	Recognize the purpose of a magnification in an image
PL	C_09	Interpret information to provide a full explanation
PL	C_15	Make a straightforward inference to recognize an explanation
PL	C_16	Retrieve and reproduce an explicitly stated detail
P/PL	E_04	Make a straightforward inference about an event
PL	H_08	Make a straightforward inference to recognize an explanation
P	W_06	Locate and recognize an explicitly stated detail
P	T_01	Recognize the main idea of a specified section of the text
PL	H_10	Recognize the reason for an author's use of simile
PL	H_13	Retrieve and recognize an explicitly stated detail
PL	P_01	Make a straightforward inference about the cause of a reaction
PL	P_02	Interpret the effect of the author's word choice

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

PL	A_14	Locate a given idea and identify its section header
PL	C_08	Make a straightforward inference to recognize an explanation
PL	C_13	Make a straightforward inference to recognize an explanation
Items at High International Benchmark (550)		
P/PL	K_01	Retrieve and reproduce 2 pieces of explicitly stated information when directed to the beginning of the text
P/PL	K_03	Make straightforward inferences to recognize an explanation of a metaphor
P/PL	K_05	Locate a text box with a heading and make a straightforward inference to provide an explanation
P/PL	K_06	Locate a text box with a heading and make an inference to recognize the best explanation
P/PL	K_07	Locate 1 (of 2) specified text box with a heading and make an interpretation to provide an explanation
P/PL	K_09	Evaluate how the format and content of a diagram convey information
P/PL	K_11	Locate and distinguish information from different sections of the text to make an inference
P/PL	K_12	Interpret and integrate information across different sections to nearly complete a table (5 of 6 entries)
P	I_03	Make a straightforward inference to provide 1 (of 2) explanation
P	I_04	Integrate information to provide 1 (of 2) geographic characteristic
P	I_05	Evaluate how the format of section headers conveys information
P	I_07	Interpret and integrate information to provide a causal explanation
P	I_09	Make a straightforward inference about the purpose of an action
P	I_11	Integrate information to provide a characteristic
P	I_13	Locate and reproduce an explicitly stated detail
P	L_03	Recognize a synonym to locate and reproduce explicitly stated information
P	L_05	Recognize an explanation of a metaphor
P	L_06	Make an inference to explain that historical documents communicate ideas
P	L_08	Make a straightforward inference to identify and reproduce explicitly stated information
P	L_10	Interpret an abstract idea by providing an example
P	L_12	Evaluate textual elements and content to provide author's point of view
P/PL	E_07	Locate and reproduce 2 explicitly stated details
P/PL	E_13	Locate and reproduce 1 characteristic
P	W_01	Retrieve and reproduce 2 pieces of information from the beginning of the text
P	W_02	Locate and interpret 1 (of 2) beneficial action

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

P	W_03	Make a straightforward inference to recognize an explanation
P	W_07	Locate information to connect 1 (of 3) action to its significance
P	T_02	Locate and reproduce 2 actions that are part of a sequence of events
P	T_03	Make a straightforward inference to provide 2 explanations
P	T_04	Make a straightforward inference to provide an explanation
P	T_05	Locate and recognize an explicitly stated action that is part of a sequence of events
P	T_07	Integrate ideas to provide an explanation
P	T_08	Locate and reproduce an explicitly stated detail
P	T_10	Locate and reproduce an explicitly stated explanation
P	T_11	Distinguish and integrate information from across different sections to nearly complete a table (4 of 5 entries)
P	T_12	Make a straightforward inference about an event
P	T_14	Evaluate the content of a diagram and interpret its meaning
P/PL	K_08	Distinguish relevant information to make an inference about a scientific explanation
P	I_10	Recognize the meaning conveyed by an image
P	I_12	Distinguish relevant information to recognize an explicitly stated reason
P	T_16	Evaluate the headings of different sections and show understanding of how the sections are divided
P	L_11	Evaluate content and generalize to recognize the most appropriate title
P/PL	E_03	Make an inference to recognize the reason for a situation
P/PL	E_09	Evaluate how the use of an image conveys information
P/PL	E_17	Integrate ideas across text to determine the main idea
P	W_08	Make an inference to recognize the purpose for an action
Items at Advanced International Benchmark (625)		
P/PL	K_07	Locate 2 specified text boxes with headings and make interpretations to provide an explanation for each
P/PL	K_10	Integrate information from 3 text boxes to provide a sequence, or use information from fewer text boxes with supporting explanation
P/PL	K_10	Integrate information from 3 text boxes with headings to provide a sequence with supporting explanation
P/PL	K_12	Interpret and integrate information across different sections to fully complete a table (5 of 6 entries)
P	I_03	Make a straightforward inference to provide 2 explanations
P	I_07	Interpret and integrate information to provide 2 causal explanations
P	I_11	Interpret and integrate information to identify a characteristic and link it to its effect

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

P	I_14	Evaluate textual elements and content to show how they exemplify the writer's point of view
P	I_15	Interpret and integrate information to provide a causal explanation
P	L_04	Locate and reproduce 2 pieces of explicitly stated information and explain the significance of 1 piece of information
P	L_10	Interpret an abstract idea by providing an example and explaining why it illustrates the abstract idea
P	W_02	Locate and interpret 2 mutually beneficial actions
P	W_07	Locate and integrate information to connect 2 actions (of 3) to their significance
P	W_07	Locate and integrate information to connect 3 actions to their significance in a sequence
P	W_11	Locate and interpret relevant information in the context of the whole text
P	W_12	Locate and interpret information to recognize the reason for a situation
P	W_13	Evaluate ideas and information across the text to make a prediction
P	T_09	Distinguish relevant information to make an inference about an action
P	T_11	Distinguish and integrate information from across different sections to fully complete a table (5 of 5 entries)
P/PL	K_04	Locate and distinguish relevant information from among several text boxes
P	I_02	Make an inference about the reason for an action
P	I_06	Make an inference about the reason for a situation
P	L_02	Distinguish relevant information across several parts of a text to recognize a possible causal inference
P	L_07	Integrate information across several parts of text to infer and recognize an explanation
P	W_05	Distinguish and recognize a paraphrase from the end of a specified text box
P	W_09	Recognize the main message of a short narrative from a specified part of the text
P	W_10	Make an inference about the reason for an action
P	T_13	Distinguish relevant information and make an inference about a scientific question
P	T_15	Evaluate textual elements and content to recognize how they exemplify the writer's point of view
Items Above the Advanced International Benchmark (625)		
P	I_04	Interpret and integrate information to provide 2 geographic characteristics
P	L_04	Locate and reproduce 2 pieces of explicitly stated information and explain the significance of both pieces of information
P	L_08	Make a straightforward inference to identify and reproduce explicitly stated information and connect this information to a later part of the text
P	L_12	Evaluate textual elements and content to provide author's point of view and support with evidence from the text
P	T_07	Integrate ideas to provide 2 explanations

P: PIRLS passage **PL: PIRLS Literacy passage** **P/PL: Passage in PIRLS and PIRLS Literacy**

Appendix 13D: ePIRLS 2016 Item Descriptions Developed During the PIRLS 2016 Benchmarking

Items at Low International Benchmark (400)		
E	M_02	Retrieve and reproduce 3 pieces of explicitly stated information from the text or the diagram
E	M_09	Retrieve and reproduce the definition of a term from a pop-up text box
E	R_02	Retrieve and reproduce explicitly stated information
E	R_06	Check the contents of 3 pop up boxes to locate and reproduce an explicitly stated detail
E	R_10	Retrieve an explicitly stated detail embedded in continuous text
E	B_01	Make a straightforward inference from a list of Internet search results to recognize the most relevant website
E	B_02	Locate and recognize explicitly stated information in a timeline
E	Z_01	Make a straightforward inference from a list of Internet search results to recognize the most relevant website
E	Z_11	Locate and recognize an explicitly stated detail
E	Z_14	Make a straightforward inference to provide 1 (of 2) aspect of a situation
E	T_02	Make a straightforward inference about a reason
E	M_05	Locate and recognize an explicitly stated reason
E	R_13	Make a straightforward inference from a list of Internet search results to recognize the most relevant website
E	Z_08	Locate and recognize an explicitly stated reason
Items at Intermediate International Benchmark (475)		
E	M_03	Make a straightforward inference to provide a reason
E	M_08	Locate and reproduce an explicitly stated reason
E	M_11	Integrate complex information in text and an animated graphic to provide a partial explanation
E	M_15	Make a straightforward inference to provide a reason
E	R_03	Integrate information from a web page to recognize 3 (of 4) connections
E	R_05	Evaluate the use of a map with interactive features to convey information
E	B_04	Locate and reproduce explicitly stated information by scrolling through a timeline
E	B_06	Make a straightforward inference about an opinion
E	B_08	Locate and reproduce an explicitly stated reaction
E	B_09	Make a straightforward inference to provide a reason
E	Z_04	Locate and recognize a reason for an action

E: ePIRLS task

E	Z_20	Integrate evidence from the text to match 3 (of 4) defense strategies with the animal(s) that uses it
E	T_16	Retrieve explicitly stated information by navigating to a labeled section of an interactive diagram
E	R_12	Make a straightforward inference to recognize a reason
E	Z_03	Evaluate the use of fact boxes containing both text and images to convey information
E	Z_05	Evaluate the use of an animated graphic to convey information
E	Z_07	Make a straightforward inference to recognize an action
E	T_04	Interpret and integrate events to recognize the cause of an outcome
E	T_11	Locate and recognize an explicitly stated detail
E	R_01	Make a straightforward inference from a list of Internet search results to recognize the most relevant website
E	Z_13	Locate and recognize explicitly stated information embedded in continuous text
E	T_12	Locate and recognize information from a map
Items at High International Benchmark (550)		
E	M_04	Locate and reproduce an explicitly stated scientific detail embedded in text
E	M_16	Interpret and integrate textual and visual information from a web page to recognize 3 (of 4) functions by navigating across interactive images
E	M_16	Interpret and integrate textual and visual information from a web page to recognize 4 functions by navigating across interactive images
E	M_17	Evaluate the writer's use of a comparison by providing 1(of 2) specific example
E	R_07	Interpret and integrate information across a web page to recognize 3 (of 4) characteristics
E	R_09	Integrate information to provide an explanation
E	R_11	Make a straightforward inference to provide 1 (of 2) piece of supporting evidence
E	R_14	Evaluate content to draw a conclusion and support it with evidence
E	R_15	Locate and reproduce 2 pieces of explicitly stated information
E	R_16	Integrate information from multiple web pages to provide a causal outcome
E	B_13	Interpret and integrate information to draw a conclusion and support it with evidence
E	B_15	Locate and compare information to provide 1 (of 2) similarity
E	B_16	Interpret and integrate information to provide 3 actions
E	B_17	Interpret and integrate information from multiple web pages to provide 2 (of 3) achievements
E	Z_02	Locate and reproduce 1 (of 2) explicitly stated similarity
E	Z_06	Make a straightforward inference to provide a prediction
E	Z_10	Evaluate the author's word choice to recognize its meaning

E: ePIRLS task

E	Z_12	Interpret and integrate visual and textual information across web pages to provide a contrast
E	Z_14	Make an inference to provide 2 contrasting aspects of a situation
E	Z_19	Interpret information to provide 1 (of 2) explanation
E	T_03	Evaluate the text to recognize how the author conveys meaning through repetition
E	T_05	Interpret and integrate information from across a web page to provide contrasting views of an event
E	T_06	Make an inference to provide support for a claim
E	T_08	Make a straightforward inference to provide a comparison
E	T_10	Make a straightforward inference about a reason
E	T_17	Evaluate how the design of an interactive diagram supports content
E	T_01	Make an inference from a list of Internet search results to distinguish the most relevant website
E	M_10	Evaluate the use of an animated diagram to determine its purpose
E	M_12	Make a straightforward inference to recognize a definition from text and images
E	B_07	Evaluate the use of punctuation to convey meaning
E	Z_15	Locate and recognize an explicitly stated detail by navigating to a pop-up box
E	Z_18	Make an inference from a list of Internet search results to distinguish the most relevant website
Items at Advanced International Benchmark (625)		
E	M_07	Evaluate textual elements to recognize the meaning of a phrase
E	M_13	Integrate information from a web page to provide an explanation
E	M_18	Make an inference to provide an explanation
E	M_20	Evaluate textual elements and content to show how they exemplify the writer's point of view
E	R_03	Integrate information from a web page to recognize 4 connections
E	R_07	Interpret and integrate information across a web page to recognize 4 characteristics
E	R_08	Evaluate the purpose of the structure of a visual display of information
E	R_11	Make a straightforward inference to provide 2 pieces of supporting evidence
E	B_05	Evaluate the use of a timeline to convey information
E	B_06	Make inferences about the opinion of two groups of people
E	B_12	Interpret and integrate information to provide a cause for an outcome
E	B_14	Locate and reproduce textual evidence to support an inference
E	Z_16	Integrate information by navigating to 2 pop-up boxes to compare and contrast actions

E: ePIRLS task

E	Z_17	Evaluate the substantive contribution of words relative to images across pages of a website
E	Z_21	Integrate information from across a web page to compare 4 actions
E	T_14	Evaluate language choices to show how they exemplify the writer's point of view
E	T_18	Integrate information from multiple web pages to order events chronologically
E	M_06	Make an inference from a list of Internet search results to distinguish the most relevant website
E	R_04	Evaluate a web page to recognize why the title fits the content
E	T_13	Interpret an integrate information to recognize how actions exemplify a principle
E	T_15	Integrate information to recognize a fact
Items Above the Advanced International Benchmark (625)		
E	M_01	Make an inference from a list of Internet search results to distinguish the most relevant website
E	M_11	Integrate complex information in text and an animated graphic to provide an explanation
E	M_14	Integrate information from across multiple web pages to provide 3 objects matched to their functions
E	M_17	Evaluate the writer's use of a comparison by providing 2 specific examples
E	M_19	Evaluate an article to determine the meaning of its title
E	B_03	Locate explicitly stated information by navigating to a pop up box via a hyperlink
E	B_10	Integrate information from multiple web pages to provide 2 actions
E	B_11	Evaluate the author's description of a family to determine her reason for the description
E	B_15	Locate and compare information to provide 2 similarities
E	B_17	Interpret and integrate information from multiple web pages to provide 3 achievements
E	Z_02	Integrate information from across a web page to compare 3 (of 4) actions
E	Z_09	Make a straightforward inference about the information provided in an animated graphic
E	Z_19	Interpret information to provide 2 explanations
E	T_07	Make a straightforward inference to identify an example of a defined term
E	T_09	Integrate information from the text to explain a phrase from the text

E: ePIRLS task

CHAPTER 14

Creating and Interpreting the PIRLS 2016 Context Questionnaire Scales

Michael O. Martin
Ina V.S. Mullis
Martin Hooper
Liqun Yin
Pierre Foy
Bethany Fishbein
Jenny Liu

Overview

As described in Chapter 2: Developing the PIRLS 2016 Context Questionnaires, many of the PIRLS 2016 context questionnaire items were developed to be combined into scales measuring a single underlying latent construct. For reporting, the scales were constructed using item response theory (IRT) scaling methods, specifically the Rasch partial credit model. As a parallel to the PIRLS International Benchmarks of achievement, each context scale allowed students to be classified into regions corresponding to high, middle, and low values on the construct. To facilitate interpretation of the regions, the cutpoints delimiting the regions were defined in terms of combinations of response categories. For certain scales that maintained many of the same items across PIRLS 2011 and PIRLS 2016, the scales were linked to allow for trend measurement on the background construct.

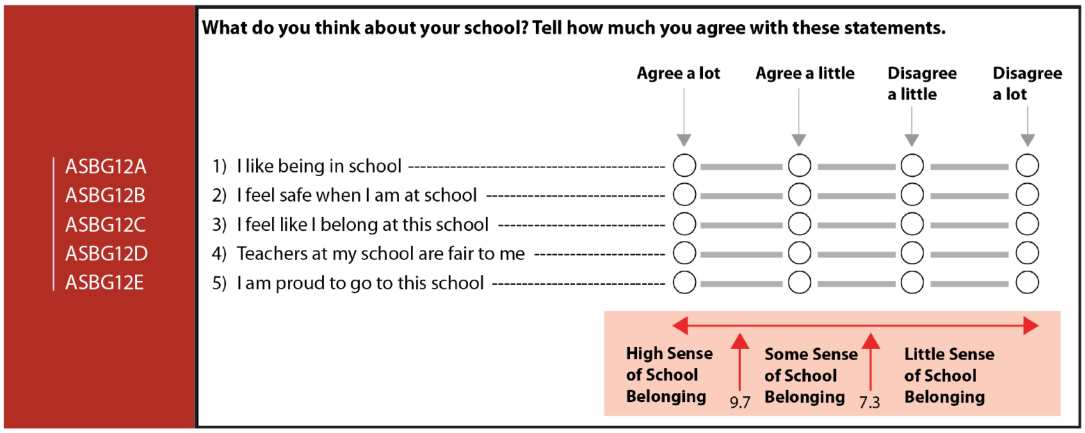
This chapter describes the procedures for constructing, interpreting, and validating scales based on responses to student, teacher, school, and home questionnaires for PIRLS and ePIRLS 2016, and then details the process for linking and reporting trend scales.

Reporting PIRLS 2016 Context Questionnaire Scales

As an example illustrating the PIRLS approach to reporting context questionnaire data, Exhibit 14.1 presents the PIRLS 2016 *Sense of School Belonging* scale. As the name suggests, this scale seeks to measure students' feelings towards their school and connectedness with the school community.

For each of the five statements, students were asked to indicate the degree of their agreement with the statement: agree a lot, agree a little, disagree a little, or disagree a lot. Using IRT partial credit scaling, the data from student responses were placed on a scale constructed so that the scale centerpoint of 10 was located at the mean score across all PIRLS countries. The units of the scale were chosen so that 2 scale score points corresponded to the logit standard deviation across all countries. Students with a **High Sense of School Belonging** had a scale score greater than or equal to the point (9.7) on the scale, corresponding to agreeing a lot, on average, with three of the five statements and agreeing a little with two of the statements. Students with **Little Sense of School Belonging** had a score no higher than the point (7.3) on the scale corresponding to disagreeing a little with three of the statements, on average, and agreeing a little with two of them.

Exhibit 14.1: Items in the PIRLS 2016 *Sense of School Belonging* Scale



Scaling Procedure

Partial credit IRT scaling is based on a statistical model that relates the probability that a person will choose a particular response to an item to that person's location on the underlying construct. In the PIRLS 2016 *Sense of School Belonging* scale, the underlying construct is students' feelings about their school, and students who agree in general with the five statements are assumed to have a greater sense of belonging, and students who disagree with the statements are assumed to feel less belonging.

The partial credit model (Masters, 1982) is shown below:

$$P_{x_i}(\theta_n) = \frac{e^{\sum_{j=0}^{x_i}\left[\theta_n - \left(\delta_i + \tau_{ij}\right)\right]}}{\sum_{h=0}^{m_i} e^{\sum_{j=0}^{h}\left[\theta_n - \left(\delta_i + \tau_{ij}\right)\right]}} \quad x_i = 0, 1, \ldots, m_i \tag{14.1}$$

$P_{x_i}(\theta_n)$ denotes the probability that person n with location θ_n on the latent construct would choose response level x_i to item i out of the m_i possible response levels for the item. The item parameter δ_i gives the location of the item on the latent construct and τ_{ij} denotes step parameters for the response levels. For each scale, the scaling procedure involves first estimating the δ_i and τ_{ij} item parameters, and then using the model with these parameters to estimate θ_n, the score on the latent construct, for each on the n respondents. Depending on the scale, respondents may be students, parents, teachers, or school principals.

The PIRLS 2016 context questionnaire scaling was conducted using the ConQuest 2.0 software (Wu, Adams, Wilson, & Haldane, 2007).

In preparation for the context questionnaire scaling effort, the TIMSS & PIRLS International Study Center developed a system of production programs that could effectively calibrate the items on each scale using ConQuest and produce scale scores for each scale respondent. The PIRLS assessment population consisted of approximately 300,000 students, as well as their parents, teachers, and school principals. The estimation of the item parameters, a procedure also known as item calibration, was conducted on the combined data from all countries, with each country contributing equally to the calibration. This was achieved by assigning weights that sum to 500 for each country's student data. Exhibit 14.2 shows the international item parameters for the *Sense of School Belonging* scale. For each item, the delta parameter δ_i shows the estimated overall location of the item on the scale, and the tau parameters τ_{ij} show the location of the steps, expressed as deviations from delta. Also, included in the right column is the Rasch infit item statistic, which is a measure of how well the data matches the model, with values above 1.3 indicating unexpected response patterns. As can be seen in this exhibit, the data seemed to match the model well for the five items in the *Sense of School Belonging* scale.

Exhibit 14.2: Item Parameters for the PIRLS 2016 *Sense of School Belonging* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ASBG12A	0.35879	-0.45320	-0.71326	1.16646	1.03
ASBG12B	-0.05809	-0.47509	-0.46271	0.93780	1.02
ASBG12C	0.05363	-0.33690	-0.38143	0.71833	1.00
ASBG12D	-0.20416	-0.29014	-0.40311	0.69325	1.10
ASBG12E	-0.15017	-0.09998	-0.47379	0.57377	0.95

Once the calibration was completed and international item parameters were estimated, individual scores for each respondent (students, teachers, principals, or parents) were generated using weighted maximum likelihood estimation (Warm, 1989). All cases with valid responses to at least two items on a scale were included in the calibration and scoring processes.

The scale scores produced by the weighted likelihood estimation are in the logit metric with measured values ranging from approximately –5 to +5. To convert to a more convenient reporting metric, a linear transformation was applied to the international distribution of logit scores to place the data from student responses on a scale constructed so that the scale centerpoint of 10 was located at the mean logit score across all TIMSS countries and 2 scale score points corresponded to the standard deviation of the logit scores across all countries. Exhibit 14.3 presents the scale transformation constants applied to the international distribution of logit scores for the *Sense of School Belonging* scale to transform them to the (10, 2) reporting metric.

This scaling approach was followed for all scales, including most of the scales in Chapter 3 of the *ePIRLS 2016 International Results in Online Informational Reading* report. The exception is the *Self-Efficacy for Computer Use* scale, the results for which are shown in Exhibit 3.5 of the ePIRLS report. This scale was composed of items included in the short ePIRLS questionnaire and for this reason the scaling of *Self-Efficacy for Computer Use* was only based on data from ePIRLS countries.

Exhibit 14.3: Scale Transformation Constants for the PIRLS 2016 *Sense of School Belonging* Scale

Scale Transformation Constants	
A = 7.558990	Transformed Scale Score = 7.558990 + 1.566579 • Logit Scale Score
B = 1.566579	

To provide an approach to reporting the context questionnaire scales analogous to the PIRLS International Benchmarks for the PIRLS achievement scales, a method was developed to divide each scale into high, middle, and low regions and provide a content-referenced interpretation for these regions. For the PIRLS achievement scales, the Low, Intermediate, High, and Advanced International Benchmarks are specific reference points on the scale that can be used to monitor progress in student achievement. Using a scale anchoring procedure, student performance at each Benchmark is described in terms of what students reaching that Benchmark know and can do. The percentage of students reaching each of these International Benchmarks can serve as a profile of student achievement in a country.

For the high, middle, and low regions of the context questionnaire scales, the interpretation is content-referenced to the extent that the boundaries of the regions were defined in terms of identifiable combinations of response categories. The particular response combinations that defined the regions boundaries, or cutpoints, were based on a judgment by PIRLS staff of what constituted a high or low region on each individual scale. For example, based on a consideration of the questions making up the *Sense of School Belonging* scale, it was determined that in order to be in the high region of the scale and labeled "High Sense of School Belonging," a student would have to agree a lot, on average, to at least three of the five statements and agree a little to the other two. Similarly, it was determined that a student who, on average, at most agreed a little with two of the statements and disagreed a little with the other three would be labeled to have "Little Sense of School Belonging."

The scale region cutpoints were quantified by assigning a numeric value to each response category, such that each respondent's responses to the scale's questions could be expressed as a "raw score." Assigning 0 to "Disagree a lot," 1 to "Disagree a little," 2 to "Agree a little," and 3 to "Agree a lot," results in raw scores on the *Sense of School Belonging* scale ranging from 0 (disagree a lot with all five statements) to 15 (agree a lot to all five). A student who agreed a lot with three statements and agreed a little with the other two would have a raw score of 13 ($3 \times 3 + 2 \times 2$). Following this approach, a student with a raw score of 13 or more would be in the "High Sense of School Belonging" region of the scale. Similarly, agreeing a little with two statements and disagreeing a little with three statements would result in a raw score of 7 ($2 \times 2 + 3 \times 1$), so that a student with a raw score less than or equal to 7 would be in the "Little Sense of School Belonging" region.

A property of a Rasch scale is that each raw score has a unique scale score associated with it. Exhibit 14.4 presents a raw score-scale score equivalence table for the *Sense of School Belonging* scale. From this table, it can be seen that a raw score of 7 corresponds to a scale score of 7.3 (rounding up) and a raw score of 13 corresponds to a scale score of 9.7 (rounding down).[1] These scale scores were the cutpoints used to divide the scale into the three regions.

Exhibit 14.4: Equivalence Table of Raw and Transformed Scale Scores for the PIRLS 2016 *Sense of School Belonging* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.45100	
1	4.84737	
2	5.50192	
3	5.96197	
4	6.33045	
5	6.65432	
6	6.95548	
7	7.25277	7.3
8	7.54290	
9	7.85416	
10	8.19786	
11	8.59455	
12	9.08205	
13	9.73132	9.7
14	10.70304	
15	12.65139	

1 The reason for rounding was to facilitate reporting, and it was decided that the highest cutpoint would be rounded down to ensure that those with an unrounded scale score (e.g., 9.73132 for the *Sense of School Belonging* scale) at the cutpoint were included within the highest region. For a similar reason, the lower cutpoint was rounded up.

Linking Procedures for Trend Context Questionnaire Scales

As a new initiative, trend results in the form of changes from 2011 to 2016 were reported for 8 PIRLS context questionnaire scales. For these trend scales, linking procedures were implemented to place the data from the two cycles on a common metric. This section describes the procedures for measuring trends—placing data for the PIRLS 2016 context questionnaire scales onto the PIRLS 2011 metric and validating this process.

As described in Chapter 2, with each cycle of PIRLS, the questionnaires are revised to keep up with the times and to improve the measurement of the constructs. Using context questionnaire IRT scales to measure background constructs began with PIRLS 2011, and during the development phase of the PIRLS 2016 questionnaires, a conscious effort was made to increase the number of items contributing to each scale in order to enhance scale reliability and validity. The context scales used to measure trends in PIRLS 2016 have items common to both PIRLS 2011 and PIRLS 2016—also called trend items—and new items unique to PIRLS 2016.

Generally, a context questionnaire scale was considered for trend reporting in 2016 if it had a sufficient number of items in common with 2011: a minimum of five common items and more than half of the PIRLS 2016 items being common items. Before deciding to measure trend on these scales, staff at the TIMSS & PIRLS International Study Center conducted extensive analysis to examine item behavior in both cycles. For example, staff at the TIMSS & PIRLS International Study Center examined differences in parameter estimates across cycles (as shown in Exhibit 14.7). Trend was only reported on those scales that appeared to have similar measurement properties across the two cycles.

As an example, Exhibit 14.5 shows the PIRLS 2016 *Parents Like Reading* scale—one of the scales where trend measurement was reported. This scale measures how students' parents feel about reading, in terms of their level of agreement with eight statements about liking reading as well as how often they read for enjoyment. Statements expressing negative sentiment were reverse coded during the scaling. Eight of the nine items were common to the PIRLS 2011 and PIRLS 2016 versions of this scale, with "T" for trend identifying these items to the left of their variable name. One new statement was added to the eight common items to improve the measure of *Parents Like Reading* for PIRLS 2016.

Exhibit 14.5: Items in the PIRLS 2016 *Parents Like Reading* Trend Scale

		Please indicate how much you agree with the following statements about reading.	Agree a lot	Agree a little	Disagree a little	Disagree a lot
T	ASBH12A*	1) I read only if I have to*	○	○	○	○
T	ASBH12B	2) I like talking about what I read with other people	○	○	○	○
T	ASBH12C	3) I like to spend my spare time reading	○	○	○	○
T	ASBH12D*	4) I read only if I need information*	○	○	○	○
T	ASBH12E	5) Reading is an important activity in my home	○	○	○	○
T	ASBH12F	6) I would like to have more time for reading	○	○	○	○
T	ASBH12G	7) I enjoy reading	○	○	○	○
	ASBH12H	8) Reading is one of my favorite hobbies	○	○	○	○

Very Much Like | 10.5 | Somewhat Like | 8.1 | Do Not Like

			Every day or almost every day	Once or twice a week	Once or twice a month	Never or almost never
T	ASBH11	**When you are at home, how often do you read for your enjoyment?**	○	○	○	○

Very Much Like | 10.5 | Somewhat Like | 8.1 | Do Not Like

*reverse coded

T Trend item—item was included in the same scale in PIRLS 2011 and was used for linking the PIRLS 2011 and PIRLS 2016 scales.

The IRT calibration and scoring methods for trend scales were the same as those used for the new context scales. The data for these nine items were calibrated across all PIRLS 2016 countries using the Rasch partial credit model, and, through this calibration, item parameters were estimated on a logit scale that was unique to the 2016 cycle. Following calibration, weighted maximum likelihood estimation was used to derive Rasch logit scale scores based on these estimated item parameters for all countries and benchmarking participants, and as such student scores were placed on this 2016 logit metric. Although similar, the PIRLS 2016 logit metric is not identical to the PIRLS 2011 logit metric, and thus the PIRLS 2016 scores needed to be transformed to the 2011 metric to allow for trend reporting.

This linking was achieved through a two-step transformation process. The first transformation—with linear constants A_1 and B_1—placed the PIRLS 2016 logit scale scores on the PIRLS 2011 logit metric, and the second transformation—with linear constants A_2 and B_2—transformed the PIRLS 2011 logit metric to the PIRLS achievement scale, which uses the (10, 2) metric described earlier. To increase the efficiency of this transformation process and reduce rounding errors, both transformations were combined into one calculation using the equations below to create a set of final scale transformation constants, A and B:

$$B = B_2 \cdot B_1 \tag{14.2}$$

$$A = A_2 + B_2 \cdot A_1 \tag{14.3}$$

The first set of transformation parameters, A_1 and B_1, were obtained by applying the mean/sigma method (Kolen & Brennan, 2004) to the two sets of common item parameters: one from the current calibration of PIRLS 2016 data and the other from the previous calibration of PIRLS 2011 data. The mean and standard deviation of the estimates of the difference between item location and item step parameter, $(\delta_i - \tau_{ij})$, were first found over all common items and all categories for each calibration. The transformation parameters A_1 and B_1 were calculated based on these two sets of means and standard deviations:

$$B_1 = \frac{SD_{c11}}{SD_{c16}} \tag{14.4}$$

$$A_1 = MN_{c11} - \frac{SD_{c11}}{SD_{c16}} \cdot MN_{c16} \tag{14.5}$$

where MN_{c16} and SD_{c16} are the mean and standard deviation of the estimates of $(\delta_i - \tau_{ij})$ of all common items and categories from the current calibration on PIRLS 2016 data; MN_{c11} and SD_{c11} are the mean and standard deviation of the estimates of $(\delta_i - \tau_{ij})$ of all common items and categories from the previous calibration on PIRLS 2011 data.

The second set of transformation parameters, A_2 and B_2, were retrieved from the scale transformations which were established in 2011 for reporting. This transformation aimed to place the resulting Rasch scores on the PIRLS (10, 2) reporting scale.

Exhibit 14.6 presents the final trend scale transformation constants applied to the PIRLS 2016 international distribution of logit scale scores for the *Parents Like Reading* trend scale to transform them to the PIRLS (10, 2) trend reporting scale.

Exhibit 14.6: Scale Transformation Constants for the PIRLS 2016 *Parents Like Reading* Trend Scale

Scale Transformation Constants	
A = 8.166833	Transformed Scale Score = 8.166833 + 1.409138 • Logit Scale Score
B = 1.409138	

To assess the accuracy of the linking, item parameter estimates for the trend items were compared across the two cycles by examining the differences between the PIRLS 2016 item parameter estimates after being transformed to the PIRLS 2011 logit metric, and the PIRLS 2011 item parameter estimates on the 2011 logit scale. Exhibit 14.7 presents the differences between these estimates for the *Parents Like Reading* trend scale. As can be seen in the exhibit, the differences were at an acceptable level for both location and step parameters, with most deviations being less than 0.1.

Exhibit 14.7: Differences in Parameter Estimates for Common Items on the PIRLS 2011 Logit Metric, *Parents Like Reading* Trend Scale

	PIRLS 2016 Variable	PIRLS 2011 Variable	Difference in delta	Difference in tau_1	Difference in tau_2	Difference in tau_3
*	ASBH12A	ASBH13A	-0.05922	-0.10022	0.04004	0.06018
	ASBH12B	ASBH13B	-0.03808	-0.10320	0.14129	-0.03809
	ASBH12C	ASBH13C	0.04140	-0.05150	0.05081	0.00070
*	ASBH12D	ASBH13D	-0.03301	-0.06083	0.04655	0.01428
	ASBH12E	ASBH13E	0.03086	-0.07248	0.10420	-0.03172
	ASBH12F	ASBH13F	-0.05200	-0.18557	0.17096	0.01460
	ASBH12G	ASBH13G	0.03937	-0.28655	0.19478	0.09177
	ASBH11	ASBH12	0.07069	-0.05021	0.16345	-0.11323

*Reverse coded

Validating the PIRLS 2016 Context Questionnaire Scales

As evidence that the context questionnaire scales provide comparable measurement across countries, reliability coefficients were computed for each scale for every country and benchmarking participant, and a principal components analysis of the scale items was conducted. Exhibit 14.8 presents the results of this analysis for the *Parents Like Reading* scale. The Cronbach's Alpha reliability coefficients generally were at an acceptable level, with all above 0.7 and many at 0.9. The exhibit also shows the percentage of variance among the scale items accounted for by the first principal component in each country. In most cases this was acceptably high, indicating that the items could be adequately represented by a single scale. The component loadings of each questionnaire item from the principal components analysis are positive and substantial, indicating a strong correlation between each item and the scale in every country.

Exhibit 14.8: Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the PIRLS 2016 *Parents Like Reading* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBH12A*	ASBH12B	ASBH12C	ASBH12D*	ASBH12E	ASBH12F	ASBH12G	ASBH12H	ASBH11
Australia	0.89	56	0.76	0.59	0.84	0.75	0.72	0.62	0.84	0.86	0.67
Austria	0.90	57	0.72	0.57	0.87	0.76	0.79	0.63	0.85	0.84	0.70
Azerbaijan	0.83	52	-	0.55	0.77	-	0.76	0.65	0.84	0.84	0.57
Bahrain	0.81	43	0.43	0.47	0.79	0.50	0.70	0.68	0.81	0.83	0.57
Belgium (Flemish)	0.90	57	0.75	0.50	0.87	0.75	0.75	0.67	0.88	0.86	0.70
Belgium (French)	0.90	55	0.70	0.52	0.85	0.72	0.72	0.71	0.86	0.86	0.70
Bulgaria	0.91	60	0.62	0.71	0.89	0.48	0.82	0.84	0.90	0.88	0.73
Canada	0.89	54	0.73	0.56	0.84	0.73	0.70	0.64	0.84	0.85	0.69
Chile	0.87	50	0.63	0.60	0.83	0.59	0.69	0.70	0.80	0.83	0.63
Chinese Taipei	0.86	49	0.57	0.48	0.77	0.54	0.76	0.77	0.86	0.86	0.56
Czech Republic	0.90	57	0.73	0.63	0.86	0.75	0.71	0.67	0.86	0.89	0.66
Denmark	0.90	57	0.74	0.60	0.88	0.78	0.76	0.60	0.85	0.83	0.70
Egypt	0.79	55	-0.25	0.78	0.83	-0.41	0.85	0.86	0.91	0.87	0.58
England	-	-	-	-	-	-	-	-	-	-	-
Finland	0.91	58	0.73	0.57	0.87	0.76	0.82	0.65	0.86	0.87	0.67
France	0.88	51	0.69	0.53	0.82	0.70	0.65	0.60	0.82	0.84	0.71
Georgia	0.78	43	0.24	0.56	0.81	0.26	0.68	0.75	0.85	0.81	0.61
Germany	0.89	54	0.75	0.57	0.87	0.66	0.70	0.61	0.86	0.85	0.68
Hong Kong SAR	0.85	48	0.46	0.52	0.78	0.41	0.75	0.78	0.87	0.87	0.61
Hungary	0.90	56	0.70	0.60	0.84	0.71	0.74	0.74	0.83	0.86	0.65
Iran, Islamic Rep. of	0.81	44	0.39	0.55	0.79	0.28	0.71	0.78	0.82	0.82	0.60
Ireland	0.88	53	0.74	0.55	0.83	0.74	0.69	0.57	0.84	0.86	0.69
Israel	0.85	47	0.65	0.50	0.81	0.65	0.68	0.57	0.79	0.84	0.62
Italy	0.88	52	0.69	0.55	0.82	0.69	0.70	0.67	0.81	0.83	0.67
Kazakhstan	0.72	39	0.11	0.51	0.74	0.20	0.70	0.73	0.77	0.82	0.60
Kuwait	0.84	47	0.52	0.49	0.80	0.58	0.75	0.70	0.81	0.84	0.54
Latvia	0.87	51	0.68	0.37	0.84	0.72	0.69	0.68	0.84	0.85	0.61
Lithuania	0.89	54	0.68	0.56	0.85	0.65	0.75	0.70	0.85	0.86	0.64
Macao SAR	0.83	46	0.38	0.56	0.76	0.35	0.76	0.78	0.87	0.87	0.55
Malta	0.86	49	0.68	0.49	0.82	0.72	0.57	0.64	0.82	0.86	0.62
Morocco	0.84	55	-0.04	0.81	0.86	-0.19	0.86	0.82	0.90	0.88	0.69
Netherlands	0.89	56	0.81	0.54	0.87	0.76	0.61	0.55	0.87	0.87	0.74

A dash (–) indicates comparable data not available.

*Reverse coded

Exhibit 14.8: Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the PIRLS 2016 *Parents Like Reading* Scale (Continued)

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBH12A*	ASBH12B	ASBH12C	ASBH12D*	ASBH12E	ASBH12F	ASBH12G	ASBH12H	ASBH11
New Zealand	0.89	54	0.73	0.53	0.85	0.72	0.73	0.61	0.84	0.87	0.68
Northern Ireland	0.91	59	0.82	0.66	0.86	0.80	0.71	0.58	0.82	0.88	0.75
Norway (5)	0.89	53	0.73	0.52	0.85	0.74	0.74	0.60	0.83	0.83	0.66
Oman	0.78	41	0.24	0.55	0.77	0.34	0.69	0.69	0.80	0.81	0.58
Poland	0.88	53	0.70	0.61	0.83	0.71	0.75	0.62	0.81	0.80	0.67
Portugal	0.87	50	0.68	0.55	0.80	0.67	0.72	0.64	0.79	0.84	0.61
Qatar	0.81	43	0.44	0.44	0.79	0.46	0.70	0.68	0.82	0.83	0.54
Russian Federation	0.86	49	0.61	0.51	0.82	0.64	0.65	0.68	0.83	0.85	0.61
Saudi Arabia	0.82	44	0.38	0.55	0.79	0.34	0.74	0.69	0.83	0.83	0.61
Singapore	0.85	49	0.58	0.43	0.82	0.54	0.72	0.75	0.86	0.87	0.59
Slovak Republic	0.90	57	0.74	0.62	0.87	0.71	0.72	0.68	0.87	0.89	0.68
Slovenia	0.88	53	0.74	0.54	0.85	0.63	0.70	0.69	0.87	0.83	0.59
South Africa	0.75	41	0.03	0.66	0.76	0.05	0.70	0.71	0.82	0.80	0.61
Spain	0.89	54	0.71	0.59	0.83	0.73	0.67	0.67	0.83	0.85	0.70
Sweden	0.89	55	0.78	0.58	0.88	0.77	0.68	0.56	0.85	0.82	0.67
Trinidad and Tobago	0.82	44	0.63	0.53	0.80	0.63	0.60	0.45	0.79	0.83	0.57
United Arab Emirates	0.80	42	0.41	0.49	0.78	0.41	0.72	0.68	0.81	0.82	0.53
United States	-	-	-	-	-	-	-	-	-	-	-
Benchmarking Participants											
Buenos Aires, Argentina	0.84	46	0.62	0.51	0.82	0.62	0.62	0.66	0.78	0.83	0.57
Ontario, Canada	0.88	53	0.73	0.53	0.83	0.73	0.69	0.64	0.83	0.85	0.67
Quebec, Canada	0.90	55	0.74	0.63	0.85	0.72	0.70	0.62	0.84	0.85	0.70
Denmark (3)	0.89	55	0.74	0.58	0.87	0.77	0.75	0.61	0.84	0.82	0.66
Norway (4)	0.88	53	0.72	0.53	0.85	0.73	0.75	0.60	0.83	0.83	0.63
Moscow City, Russian Fed.	0.86	49	0.65	0.47	0.82	0.69	0.67	0.63	0.83	0.83	0.61
Eng/Afr/Zulu - RSA (5)	0.78	41	0.28	0.62	0.77	0.29	0.69	0.60	0.81	0.80	0.63
Andalusia, Spain	0.89	55	0.71	0.59	0.83	0.73	0.69	0.67	0.85	0.86	0.70
Madrid, Spain	0.87	51	0.69	0.55	0.82	0.74	0.68	0.58	0.81	0.85	0.65
Abu Dhabi, UAE	0.79	41	0.39	0.50	0.78	0.37	0.73	0.66	0.81	0.82	0.49
Dubai, UAE	0.82	44	0.54	0.44	0.78	0.56	0.71	0.66	0.81	0.82	0.57

A dash (–) indicates comparable data not available.

*Reverse coded

As indicators of effective environments for learning, a positive relationship with achievement is an important aspect of validity for the PIRLS context questionnaire scales. For the *Parents Like Reading* scale, Exhibit 14.9 presents the Pearson correlation with reading achievement in PIRLS 2016 for each country, together with *r*-squared—the proportion of variance in reading achievement attributable to the *Parents Like Reading* scale. These figures show a moderate relationship with achievement across participating countries. Also shown is the proportion of variance in achievement attributable to differences between the regions of the *Parents Like Reading* scale. This is very similar to the proportion of variance explained by the scale as a whole, indicating that dividing the scale into regions loses little of its power to account for achievement differences.

Exhibit 14.9: Relationship Between the PIRLS 2016 *Parents Like Reading* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.24	0.06	0.05
Austria	0.30	0.09	0.08
Azerbaijan	0.11	0.01	0.01
Bahrain	0.19	0.04	0.03
Belgium (Flemish)	0.23	0.05	0.05
Belgium (French)	0.28	0.08	0.08
Bulgaria	0.41	0.17	0.16
Canada	0.21	0.04	0.04
Chile	0.23	0.05	0.05
Chinese Taipei	0.19	0.04	0.04
Czech Republic	0.28	0.08	0.07
Denmark	0.21	0.04	0.04
Egypt	0.30	0.09	0.08
England	-	-	-
Finland	0.25	0.06	0.06
France	0.25	0.06	0.05
Georgia	0.20	0.04	0.04
Germany	0.35	0.12	0.11
Hong Kong SAR	0.09	0.01	0.01
Hungary	0.35	0.12	0.11
Iran, Islamic Rep. of	0.25	0.06	0.05
Ireland	0.26	0.07	0.05
Israel	0.18	0.03	0.03
Italy	0.22	0.05	0.04
Kazakhstan	0.09	0.01	0.01
Kuwait	0.17	0.03	0.02
Latvia	0.22	0.05	0.05
Lithuania	0.24	0.06	0.05
Macao SAR	0.13	0.02	0.01
Malta	0.14	0.02	0.02
Morocco	0.24	0.06	0.05
Netherlands	0.26	0.07	0.06

A dash (–) indicates comparable data not available.

Exhibit 14.9: Relationship Between the PIRLS 2016 *Parents Like Reading* Scale and PIRLS 2016 Reading Achievement (Continued)

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
New Zealand	0.29	0.09	0.07
Northern Ireland	0.18	0.03	0.03
Norway (5)	0.25	0.06	0.05
Oman	0.20	0.04	0.03
Poland	0.21	0.05	0.04
Portugal	0.21	0.05	0.04
Qatar	0.21	0.05	0.04
Russian Federation	0.22	0.05	0.04
Saudi Arabia	0.14	0.02	0.02
Singapore	0.20	0.04	0.04
Slovak Republic	0.38	0.15	0.11
Slovenia	0.26	0.07	0.06
South Africa	0.17	0.03	0.03
Spain	0.20	0.04	0.04
Sweden	0.26	0.07	0.06
Trinidad and Tobago	0.12	0.02	0.02
United Arab Emirates	0.22	0.05	0.04
United States	-	-	-
International Median	**0.22**	**0.05**	**0.04**
Benchmarking Participants			
Buenos Aires, Argentina	0.24	0.06	0.06
Ontario, Canada	0.20	0.04	0.04
Quebec, Canada	0.17	0.03	0.03
Denmark (3)	0.18	0.03	0.03
Norway (4)	0.24	0.06	0.05
Moscow City, Russian Fed.	0.21	0.04	0.04
Eng/Afr/Zulu - RSA (5)	0.16	0.02	0.03
Andalusia, Spain	0.22	0.05	0.05
Madrid, Spain	0.18	0.03	0.03
Abu Dhabi, UAE	0.22	0.05	0.04
Dubai, UAE	0.23	0.05	0.05

A dash (–) indicates comparable data not available.

Item parameter estimates and item and scale statistics similar to those above are available in Appendix 14A for each of the PIRLS 2016 context questionnaire scales and in Appendix 14B for one context questionnaire scale based on responses to a brief questionnaire completed by students who participated in ePIRLS.

Reporting the PIRLS 2016 Trend Context Questionnaire Scales

Exhibit 14.10 shows an excerpt from the *Parents Like Reading* exhibit. To represent trends from 2011, the two columns to the right of the exhibit present the average scale score in 2016 for each country and the difference from the average in 2011, respectively. Up and down arrows indicate whether the trend difference is significantly higher or lower in 2016, with a 99% level of confidence.

Trend results were not reported for the percentage of students in each region. To facilitate interpretation of the region boundaries in terms of combinations of response categories, trend scales followed the same procedure as non-trend scales in setting cutpoints for classification into regions. As such, the procedure was primarily dependent on similarities in response patterns without taking into account variations in difficulty across the items that were unique to 2011 or 2016. Consequently, although the cutpoints generally are quite close across the two cycles, they are not identical and therefore it was considered most appropriate to use differences in scale score means rather than changes in the percentages in scale regions as indicators of trend.

Exhibit 14.10: Excerpt from the PIRLS 2016 *Parents Like Reading* Exhibit

Students Categorized by Parents' Reports

Students were scored on the *Parents Like Reading* scale according to their parents' responses to eight statements about reading as well as how often they read for enjoyment. Students whose parents **Very Much Like** reading had a score on the scale of at least 10.5, which corresponds to their parents "agreeing a lot" with four of the eight statements and "agreeing a little" with the other four, as well as reading for enjoyment "every day or almost every day," on average. Students whose parents **Do Not Like** reading had a score no higher than 8.1, which corresponds to their parents "disagreeing a little" with four of the eight statements and "agreeing a little" with the other four, as well as reading for enjoyment only "once or twice a month," on average. All other students had parents who **Somewhat Like** reading.

Country		Very Much Like		Somewhat Like		Do Not Like		Average Scale Score		Difference in Average Scale Score from 2011	
		Percent of Students	Average Achievement	Percent of Students	Average Achievement	Percent of Students	Average Achievement				
Ireland		47 (1.1)	588 (2.9)	40 (1.0)	560 (2.9)	13 (0.7)	544 (5.2)	10.3 (0.05)		-0.4 (0.07)	▼
Netherlands	s	46 (1.5)	566 (2.8)	39 (1.2)	548 (3.1)	16 (0.9)	525 (4.3)	10.0 (0.05)	s	-0.4 (0.08)	▼
Malta		45 (0.8)	471 (2.3)	42 (0.9)	451 (2.6)	13 (0.5)	439 (3.9)	10.2 (0.03)		-0.4 (0.05)	▼
Denmark		44 (1.1)	564 (2.3)	38 (1.0)	542 (2.7)	17 (0.7)	530 (3.6)	10.1 (0.05)		-0.6 (0.06)	▼
Sweden		44 (1.1)	576 (2.9)	42 (1.1)	551 (3.0)	13 (0.8)	529 (4.4)	10.1 (0.05)		-0.8 (0.07)	▼
Azerbaijan		44 (1.3)	481 (5.1)	46 (1.1)	470 (4.4)	10 (0.9)	453 (6.8)	10.4 (0.06)		0.7 (0.08)	▲
Norway (5)		42 (1.0)	574 (2.6)	44 (0.9)	556 (2.5)	15 (0.9)	532 (3.9)	10.1 (0.05)		- -	
Trinidad and Tobago		41 (1.0)	499 (3.6)	49 (0.9)	478 (4.3)	10 (0.6)	468 (6.1)	10.2 (0.04)		-0.5 (0.06)	▼
Finland		41 (1.0)	585 (2.2)	43 (0.9)	563 (2.5)	16 (0.7)	542 (3.4)	10.0 (0.05)		-0.5 (0.07)	▼

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Source: The full *Parents Like Reading* exhibit can be found within the *PIRLS 2016 International Results in Reading* report.

References

Kolen, M.J. & Brennan, R.L. (2004). *Test equating, scaling, and linking* (2nd Ed.). NY: Springer Science + Business media, Inc.

Masters, G.N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika, 47*(2), 149-174.

Warm, T.A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450.

Wu, M.L., Adams, R.J., Wilson, M.R., & Haldane, S. (2007). Conquest 2.0 [computer software]. Camberwell, Australia: Australian Council for Educational Research.

Appendix 14A: PIRLS 2016 Context Questionnaire Scales

Classroom Instruction Limited by Student Attributes Scale

The *Classroom Instruction Limited by Student Attributes* (SLI) scale was created based on teachers' responses concerning seven attributes of students described below.

Items in the PIRLS 2016 *Classroom Instruction Limited by Student Attributes* Scale

In your view, to what extent do the following limit how you teach this class?

		Not at all	Some	A lot
ATBR05A	1) Students lacking prerequisite knowledge or skills	◯	◯	◯
ATBR05B	2) Students suffering from lack of basic nutrition	◯	◯	◯
ATBR05C	3) Students suffering from not enough sleep	◯	◯	◯
ATBR05D	4) Students absent from class	◯	◯	◯
ATBR05E	5) Disruptive students	◯	◯	◯
ATBR05F	6) Uninterested students	◯	◯	◯
ATBR05G	7) Students with mental, emotional, or psychological impairment	◯	◯	◯

Very Little | 11.0 | Some | 6.2 | A Lot

Item Parameters for the PIRLS 2016 *Classroom Instruction Limited by Student Attributes* Scale

Item	delta	tau_1	tau_2	Infit
ATBR05A	0.86545	-1.95409	1.95409	1.01
ATBR05B	-1.21694	-1.05765	1.05765	1.04
ATBR05C	-0.29564	-1.60966	1.60966	0.97
ATBR05D	-0.06615	-1.48492	1.48492	1.04
ATBR05E	0.34789	-1.40914	1.40914	0.99
ATBR05F	0.48997	-1.87068	1.87068	0.91
ATBR05G	-0.12458	-1.47958	1.47958	1.05

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Classroom Instruction Limited by Student Attributes* Scale

Scale Transformation Constants	
A = 8.565173	Transformed Scale Score = 8.565173 + 1.261182 • Logit Scale Score
B = 1.261182	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Classroom Instruction Limited by Student Attributes* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.09130	
1	4.66178	
2	5.50732	
3	6.15928	6.2
4	6.73472	
5	7.28482	
6	7.84240	
7	8.42725	
8	9.04720	
9	9.68742	
10	10.33571	
11	11.00861	11.0
12	11.75725	
13	12.70534	
14	14.37342	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Classroom Instruction Limited by Student Attributes* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item						
			ATBR05A	ATBR05B	ATBR05C	ATBR05D	ATBR05E	ATBR05F	ATBR05G
Australia	0.83	49	0.68	0.66	0.76	0.65	0.71	0.72	0.72
Austria	0.77	42	0.69	0.56	0.74	0.58	0.61	0.71	0.64
Azerbaijan	0.68	35	0.66	0.44	0.51	0.67	0.49	0.72	0.58
Bahrain	0.77	43	0.64	0.60	0.62	0.45	0.73	0.84	0.67
Belgium (Flemish)	0.69	35	0.63	0.47	0.62	0.54	0.64	0.70	0.53
Belgium (French)	0.72	37	0.61	0.55	0.71	0.60	0.52	0.64	0.64
Bulgaria	0.77	42	0.69	0.46	0.54	0.70	0.65	0.78	0.68
Canada	0.75	40	0.64	0.69	0.72	0.54	0.59	0.55	0.69
Chile	0.75	41	0.58	0.68	0.62	0.68	0.69	0.74	0.40
Chinese Taipei	0.63	32	0.47	0.44	0.62	0.59	0.59	0.61	0.60
Czech Republic	0.65	32	0.64	0.10	0.42	0.53	0.67	0.69	0.66
Denmark	0.76	41	0.61	0.50	0.69	0.49	0.78	0.79	0.57
Egypt	0.66	33	0.57	0.59	0.62	0.46	0.74	0.61	0.39
England	0.76	41	0.67	0.55	0.71	0.71	0.64	0.62	0.59
Finland	0.69	36	0.53	0.49	0.71	0.44	0.65	0.71	0.58
France	0.77	42	0.60	0.58	0.67	0.69	0.70	0.63	0.64
Georgia	0.80	46	0.57	0.65	0.74	0.73	0.68	0.70	0.66
Germany	0.77	43	0.67	0.61	0.67	0.67	0.66	0.62	0.67
Hong Kong SAR	0.64	32	0.63	0.39	0.45	0.58	0.55	0.70	0.60
Hungary	0.75	40	0.66	0.68	0.71	0.50	0.53	0.65	0.69
Iran, Islamic Rep. of	0.76	41	0.50	0.54	0.71	0.69	0.69	0.58	0.73
Ireland	0.77	44	0.71	0.71	0.73	0.65	0.58	0.71	0.49
Israel	0.88	57	0.68	0.71	0.77	0.84	0.77	0.80	0.71
Italy	0.61	30	0.61	0.28	0.47	0.52	0.67	0.66	0.51
Kazakhstan	0.83	50	0.54	0.62	0.69	0.72	0.79	0.79	0.77
Kuwait	0.66	34	0.63	0.54	0.46	0.61	0.66	0.71	0.36
Latvia	0.81	47	0.59	0.71	0.69	0.74	0.72	0.68	0.66
Lithuania	0.79	45	0.69	0.48	0.76	0.60	0.70	0.68	0.75
Macao SAR	0.72	38	0.72	0.28	0.55	0.63	0.70	0.60	0.71
Malta	0.85	53	0.59	0.73	0.79	0.75	0.76	0.76	0.68
Morocco	0.70	36	0.49	0.63	0.62	0.59	0.67	0.74	0.43
Netherlands	0.80	48	0.59	0.70	0.76	0.67	0.71	0.69	0.73
New Zealand	0.76	41	0.53	0.73	0.71	0.68	0.59	0.64	0.59
Northern Ireland	0.80	45	0.68	0.61	0.76	0.61	0.57	0.73	0.72
Norway (5)	0.74	40	0.61	0.57	0.66	0.58	0.66	0.67	0.66
Oman	0.80	46	0.55	0.66	0.70	0.76	0.76	0.71	0.56
Poland	0.64	33	0.61	0.57	0.55	0.63	0.53	0.71	0.39
Portugal	0.77	43	0.52	0.64	0.67	0.73	0.75	0.69	0.57
Qatar	0.71	37	0.50	0.51	0.68	0.46	0.72	0.74	0.58
Russian Federation	0.86	54	0.54	0.67	0.76	0.80	0.77	0.79	0.77
Saudi Arabia	0.68	35	0.52	0.60	0.52	0.42	0.68	0.72	0.59
Singapore	0.82	48	0.65	0.68	0.71	0.76	0.66	0.66	0.73
Slovak Republic	0.78	44	0.76	0.55	0.61	0.72	0.68	0.74	0.57
Slovenia	0.78	44	0.62	0.60	0.71	0.41	0.78	0.76	0.69
South Africa	0.72	38	0.35	0.50	0.60	0.59	0.73	0.72	0.72
Spain	0.81	46	0.68	0.62	0.77	0.69	0.69	0.69	0.64
Sweden	0.68	35	0.54	0.58	0.57	0.54	0.63	0.71	0.52
Trinidad and Tobago	0.80	44	0.66	0.62	0.67	0.72	0.64	0.71	0.63
United Arab Emirates	0.80	46	0.58	0.71	0.68	0.63	0.70	0.75	0.70
United States	0.80	45	0.66	0.63	0.71	0.62	0.70	0.70	0.66
Benchmarking Participants									
Buenos Aires, Argentina	0.86	54	0.64	0.81	0.81	0.76	0.69	0.75	0.63
Ontario, Canada	0.75	41	0.74	0.70	0.76	0.56	0.40	0.58	0.69
Quebec, Canada	0.75	41	0.53	0.69	0.74	0.52	0.72	0.46	0.73
Denmark (3)	0.77	43	0.66	0.63	0.73	0.63	0.67	0.68	0.60
Norway (4)	0.72	38	0.52	0.67	0.68	0.49	0.64	0.67	0.61
Moscow City, Russian Fed.	0.79	45	0.55	0.51	0.64	0.68	0.79	0.72	0.76
Eng/Afr/Zulu - RSA (5)	0.64	34	0.59	0.43	0.47	0.21	0.76	0.81	0.59
Andalusia, Spain	0.77	44	0.70	0.66	0.72	0.65	0.60	0.71	0.61
Madrid, Spain	0.73	39	0.72	0.54	0.75	0.63	0.59	0.53	0.60
Abu Dhabi, UAE	0.76	41	0.57	0.72	0.56	0.59	0.71	0.76	0.57
Dubai, UAE	0.82	49	0.66	0.68	0.76	0.69	0.63	0.73	0.72

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Classroom Instruction Limited by Student Attributes* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.24	0.06	0.05
Austria	0.14	0.02	0.02
Azerbaijan	0.05	0.00	0.00
Bahrain	0.19	0.03	0.02
Belgium (Flemish)	0.21	0.04	0.03
Belgium (French)	0.24	0.06	0.03
Bulgaria	0.28	0.08	0.05
Canada	0.18	0.03	0.02
Chile	0.24	0.06	0.05
Chinese Taipei	0.03	0.00	0.00
Czech Republic	0.12	0.02	0.01
Denmark	0.10	0.01	0.01
Egypt	0.13	0.02	0.02
England	0.14	0.02	0.02
Finland	0.13	0.02	0.01
France	0.15	0.02	0.02
Georgia	0.01	0.00	0.01
Germany	0.27	0.07	0.05
Hong Kong SAR	0.15	0.02	0.02
Hungary	0.24	0.06	0.04
Iran, Islamic Rep. of	0.06	0.00	0.01
Ireland	0.21	0.05	0.02
Israel	0.27	0.07	0.07
Italy	0.03	0.00	0.00
Kazakhstan	-0.10	0.01	0.01
Kuwait	0.08	0.01	0.01
Latvia	0.01	0.00	0.01
Lithuania	0.13	0.02	0.01
Macao SAR	0.10	0.01	0.01
Malta	0.11	0.01	0.02
Morocco	0.21	0.04	0.05
Netherlands	0.18	0.03	0.04
New Zealand	0.23	0.06	0.04
Northern Ireland	0.19	0.03	0.01
Norway (5)	0.16	0.02	0.02
Oman	0.10	0.01	0.01
Poland	0.13	0.02	0.01
Portugal	0.10	0.01	0.01
Qatar	0.21	0.04	0.04
Russian Federation	0.16	0.03	0.02
Saudi Arabia	0.20	0.04	0.04
Singapore	0.37	0.13	0.11
Slovak Republic	0.26	0.07	0.09
Slovenia	0.04	0.00	0.00
South Africa	0.04	0.00	0.01
Spain	0.12	0.02	0.01
Sweden	0.11	0.01	0.01
Trinidad and Tobago	0.21	0.04	0.06
United Arab Emirates	0.35	0.12	0.09
United States	0.24	0.06	0.05
International Median	0.15	0.02	0.02
Benchmarking Participants			
Buenos Aires, Argentina	0.17	0.03	0.03
Ontario, Canada	0.20	0.04	0.03
Quebec, Canada	0.18	0.03	0.04
Denmark (3)	0.11	0.01	0.01
Norway (4)	0.07	0.00	0.00
Moscow City, Russian Fed.	0.09	0.01	0.01
Eng/Afr/Zulu - RSA (5)	0.08	0.01	0.00
Andalusia, Spain	0.19	0.04	0.03
Madrid, Spain	0.16	0.03	0.02
Abu Dhabi, UAE	0.25	0.06	0.07
Dubai, UAE	0.30	0.09	0.07

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

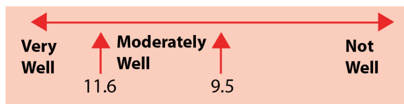
Could Do Early Literacy Tasks When Beginning Primary School Scale

The *Could Do Early Literacy Tasks When Beginning Primary School* (ELT) scale was created based on parents' responses to how well their children could do the tasks described below when they began primary school.

Items in the PIRLS 2016 *Could Do Early Literacy Tasks When Beginning Primary School* Scale

How well could your child do the following when he/she began the first grade of primary/elementary school?

			Very well	Moderately well	Not very well	Not at all
T	ASBH07A	1) Recognize most of the letters of the alphabet				
T	ASBH07B	2) Read some words				
T	ASBH07C	3) Read sentences				
	ASBH07D	4) Read a story				
T	ASBH07E	5) Write letters of the alphabet				
T	ASBH07F	6) Write some words				



T Trend item—item was included in the same scale in PIRLS 2011 and was used for linking the PIRLS 2011 and PIRLS 2016 scales.

Item Parameters for the PIRLS 2016 *Could Do Early Literacy Tasks When Beginning Primary School* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ASBH07A	-1.92907	-2.74251	-0.22290	2.96541	1.23
ASBH07B	-0.25362	-2.61266	-0.23840	2.85106	0.86
ASBH07C	1.33018	-2.27105	-0.26638	2.53743	0.87
ASBH07D	2.23418	-2.17667	-0.33003	2.50670	1.13
ASBH07E	-1.28432	-3.05497	-0.09985	3.15482	1.17
ASBH07F	-0.09735	-2.74264	-0.24088	2.98352	1.04

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Could Do Early Literacy Tasks When Beginning Primary School* Scale

Scale Transformation Constants	
A = 9.406273	Transformed Scale Score = 9.406273 + 0.729914 • Logit Scale Score
B = 0.729914	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Could Do Early Literacy Tasks When Beginning Primary School* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	4.87072	
1	5.91978	
2	6.57986	
3	7.11898	
4	7.58223	
5	8.00017	
6	8.38767	
7	8.75301	
8	9.10285	
9	9.44462	9.5
10	9.78407	
11	10.12525	
12	10.47276	
13	10.83403	
14	11.21961	
15	11.64946	11.6
16	12.15650	
17	12.80602	
18	13.88211	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Could Do Early Literacy Tasks When Beginning Primary School* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			ASBH07A	ASBH07B	ASBH07C	ASBH07D	ASBH07E	ASBH07F
Australia	0.92	71	0.75	0.89	0.89	0.84	0.82	0.88
Austria	0.90	67	0.81	0.87	0.85	0.77	0.79	0.82
Azerbaijan	0.92	70	0.79	0.89	0.87	0.77	0.84	0.86
Bahrain	0.90	69	0.76	0.89	0.88	0.79	0.78	0.86
Belgium (Flemish)	0.92	71	0.78	0.89	0.88	0.83	0.82	0.85
Belgium (French)	0.89	65	0.72	0.85	0.85	0.78	0.79	0.82
Bulgaria	0.94	75	0.82	0.92	0.89	0.84	0.85	0.89
Canada	0.91	70	0.74	0.88	0.89	0.84	0.79	0.86
Chile	0.93	73	0.79	0.88	0.90	0.85	0.85	0.86
Chinese Taipei	0.91	70	0.81	0.88	0.88	0.82	0.83	0.80
Czech Republic	0.93	73	0.79	0.90	0.90	0.85	0.82	0.87
Denmark	0.91	70	0.77	0.89	0.88	0.81	0.81	0.86
Egypt	0.93	73	0.84	0.91	0.87	0.78	0.83	0.89
England	-	-	-	-	-	-	-	-
Finland	0.92	73	0.75	0.91	0.91	0.85	0.80	0.89
France	0.87	61	0.70	0.85	0.83	0.75	0.75	0.81
Georgia	0.92	71	0.79	0.89	0.89	0.78	0.84	0.87
Germany	0.89	64	0.77	0.85	0.84	0.76	0.77	0.80
Hong Kong SAR	0.90	72	0.80	0.88	0.86	-	0.85	0.83
Hungary	0.93	73	0.82	0.90	0.90	0.80	0.86	0.84
Iran, Islamic Rep. of	0.93	75	0.82	0.89	0.90	0.83	0.87	0.89
Ireland	0.93	75	0.78	0.89	0.91	0.87	0.85	0.88
Israel	0.90	69	0.72	0.88	0.88	0.82	0.80	0.86
Italy	0.91	68	0.79	0.88	0.85	0.78	0.81	0.84
Kazakhstan	0.90	66	0.73	0.85	0.87	0.80	0.79	0.83
Kuwait	0.92	72	0.80	0.89	0.88	0.82	0.83	0.87
Latvia	0.91	69	0.76	0.87	0.90	0.83	0.78	0.84
Lithuania	0.91	69	0.77	0.88	0.89	0.84	0.77	0.83
Macao SAR	0.88	62	0.77	0.85	0.86	0.70	0.80	0.75
Malta	0.91	69	0.74	0.88	0.89	0.83	0.80	0.86
Morocco	0.94	76	0.86	0.92	0.90	0.77	0.87	0.90
Netherlands	0.92	72	0.77	0.89	0.89	0.83	0.83	0.86
New Zealand	0.90	71	0.77	0.89	0.90	0.85	0.80	-
Northern Ireland	-	-	-	-	-	-	-	-
Norway (5)	0.91	70	0.79	0.89	0.88	0.79	0.82	0.86
Oman	0.90	67	0.76	0.88	0.87	0.78	0.77	0.85
Poland	0.92	72	0.79	0.89	0.89	0.83	0.83	0.87
Portugal	0.90	67	0.73	0.88	0.87	0.80	0.79	0.83
Qatar	0.90	68	0.74	0.88	0.87	0.79	0.80	0.86
Russian Federation	0.91	70	0.77	0.90	0.90	0.84	0.78	0.83
Saudi Arabia	0.91	70	0.76	0.89	0.89	0.80	0.80	0.88
Singapore	0.92	72	0.77	0.89	0.89	0.85	0.81	0.86
Slovak Republic	0.91	69	0.78	0.89	0.87	0.77	0.81	0.83
Slovenia	0.93	73	0.81	0.90	0.89	0.82	0.84	0.87
South Africa	0.88	62	0.69	0.82	0.84	0.81	0.75	0.79
Spain	0.93	73	0.81	0.89	0.90	0.85	0.84	0.86
Sweden	0.92	72	0.80	0.89	0.89	0.81	0.83	0.88
Trinidad and Tobago	0.88	63	0.68	0.84	0.86	0.82	0.75	0.80
United Arab Emirates	0.91	69	0.77	0.88	0.88	0.81	0.79	0.86
United States	-	-	-	-	-	-	-	-
Benchmarking Participants								
Buenos Aires, Argentina	0.89	66	0.75	0.86	0.86	0.80	0.78	0.81
Ontario, Canada	0.92	72	0.76	0.89	0.90	0.85	0.81	0.86
Quebec, Canada	0.90	66	0.71	0.87	0.86	0.80	0.78	0.85
Denmark (3)	0.91	68	0.77	0.89	0.88	0.80	0.77	0.86
Norway (4)	0.91	70	0.79	0.89	0.88	0.77	0.82	0.86
Moscow City, Russian Fed.	0.91	69	0.75	0.88	0.90	0.84	0.78	0.83
Eng/Afr/Zulu - RSA (5)	0.88	63	0.68	0.83	0.85	0.82	0.75	0.81
Andalusia, Spain	0.93	75	0.83	0.90	0.91	0.85	0.85	0.87
Madrid, Spain	0.92	73	0.81	0.89	0.90	0.84	0.83	0.87
Abu Dhabi, UAE	0.91	70	0.78	0.88	0.88	0.81	0.79	0.86
Dubai, UAE	0.90	68	0.74	0.87	0.87	0.82	0.79	0.84

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Could Do Early Literacy Tasks When Beginning Primary School* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.22	0.05	0.04
Austria	-0.02	0.00	0.01
Azerbaijan	0.18	0.03	0.03
Bahrain	0.38	0.14	0.12
Belgium (Flemish)	-0.04	0.00	0.00
Belgium (French)	0.09	0.01	0.01
Bulgaria	0.44	0.19	0.13
Canada	0.30	0.09	0.08
Chile	0.31	0.10	0.08
Chinese Taipei	0.35	0.12	0.12
Czech Republic	0.16	0.03	0.03
Denmark	0.33	0.11	0.09
Egypt	0.42	0.18	0.17
England	-	-	-
Finland	0.40	0.16	0.13
France	0.19	0.04	0.03
Georgia	0.17	0.03	0.02
Germany	0.11	0.01	0.01
Hong Kong SAR	0.36	0.13	0.11
Hungary	0.10	0.01	0.01
Iran, Islamic Rep. of	0.19	0.04	0.02
Ireland	0.39	0.16	0.13
Israel	0.03	0.00	0.00
Italy	0.09	0.01	0.01
Kazakhstan	0.18	0.03	0.04
Kuwait	0.30	0.09	0.09
Latvia	0.37	0.13	0.12
Lithuania	0.45	0.20	0.18
Macao SAR	0.31	0.10	0.09
Malta	0.29	0.08	0.07
Morocco	0.35	0.12	0.13
Netherlands	0.17	0.03	0.03
New Zealand	0.13	0.02	0.02
Northern Ireland	-	-	-
Norway (5)	0.26	0.07	0.05
Oman	0.39	0.15	0.13
Poland	0.29	0.08	0.07
Portugal	0.14	0.02	0.02
Qatar	0.29	0.08	0.07
Russian Federation	0.35	0.12	0.11
Saudi Arabia	0.17	0.03	0.03
Singapore	0.46	0.21	0.19
Slovak Republic	0.15	0.02	0.01
Slovenia	0.26	0.07	0.07
South Africa	0.19	0.04	0.04
Spain	0.35	0.12	0.11
Sweden	0.38	0.15	0.12
Trinidad and Tobago	0.34	0.12	0.10
United Arab Emirates	0.37	0.13	0.13
United States	-	-	-
International Median	0.29	0.08	0.07
Benchmarking Participants			
Buenos Aires, Argentina	0.23	0.05	0.05
Ontario, Canada	0.34	0.12	0.10
Quebec, Canada	0.19	0.04	0.04
Denmark (3)	0.34	0.11	0.11
Norway (4)	0.31	0.10	0.07
Moscow City, Russian Fed.	0.39	0.15	0.14
Eng/Afr/Zulu - RSA (5)	0.21	0.05	0.05
Andalusia, Spain	0.33	0.11	0.09
Madrid, Spain	0.33	0.11	0.09
Abu Dhabi, UAE	0.41	0.17	0.16
Dubai, UAE	0.30	0.09	0.08

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Digital Devices in the Home Scale

The *Digital Devices in the Home* (DDH) scale was created based on parents' responses concerning the availability of three resources described below.

Items in the PIRLS 2016 *Digital Devices in the Home* Scale

ASDG05D[1]

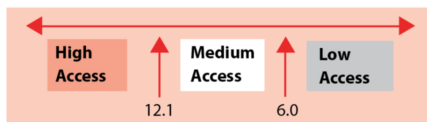
Home computer/tablet or Internet connection (students):
1) None
2) Computer/tablet or Internet connection
3) Both

ASBH16[1]

Number of digital information devices in the home (parents):
1) None
2) 1-3 devices
3) 4-6 devices
4) 7-10 devices
5) More than 10 devices

ASDH15[1]

Digital device for reading (parents):
1) None
2) Either for the parent or for the child
3) Both



1 Derived variable. For more details, see Supplement 3 of the *User Guide for the PIRLS 2016 International Database*.

Item Parameters for the PIRLS 2016 *Digital Devices in the Home* Scale

Item	delta	tau_1	tau_2	tau_3	tau_4	Infit
ASDG05D	-1.01084	0.19104	-0.19104			0.97
ASBH16	0.78901	-2.55418	-0.25050	0.94345	1.86123	1.01
ASDH15	0.22183	0.40387	-0.40387			1.04

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Digital Devices in the Home* Scale

Scale Transformation Constants	
A = 8.607915	Transformed Scale Score = 8.607915 + 1.618512 • Logit Scale Score
B = 1.618512	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Digital Devices in the Home* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.92341	
1	5.90375	6.0
2	6.97819	
3	7.91040	
4	8.78694	
5	9.68170	
6	10.78844	
7	12.19680	12.1
8	14.32363	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Digital Devices in the Home* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item		
			ASDG05D	ASBH16	ASDH15
Australia	0.32	43	0.58	0.71	0.68
Austria	0.38	45	0.48	0.78	0.72
Azerbaijan	0.77	68	0.80	0.82	0.86
Bahrain	0.38	45	0.62	0.69	0.70
Belgium (Flemish)	0.25	40	0.46	0.74	0.68
Belgium (French)	0.33	44	0.60	0.72	0.66
Bulgaria	0.58	57	0.77	0.76	0.73
Canada	0.34	44	0.58	0.71	0.69
Chile	0.61	57	0.68	0.79	0.78
Chinese Taipei	0.36	44	0.60	0.69	0.70
Czech Republic	0.35	44	0.57	0.76	0.65
Denmark	0.18	39	0.49	0.70	0.66
Egypt	0.70	62	0.74	0.79	0.84
England	-	-	-	-	-
Finland	0.21	39	0.44	0.70	0.70
France	0.30	42	0.59	0.69	0.67
Georgia	0.63	58	0.75	0.77	0.76
Germany	0.41	46	0.58	0.72	0.72
Hong Kong SAR	0.27	41	0.49	0.70	0.70
Hungary	0.50	53	0.73	0.70	0.75
Iran, Islamic Rep. of	0.73	65	0.76	0.79	0.86
Ireland	0.33	43	0.49	0.73	0.72
Israel	0.31	44	0.69	0.73	0.56
Italy	0.41	47	0.64	0.73	0.68
Kazakhstan	0.65	59	0.76	0.74	0.81
Kuwait	0.31	43	0.48	0.73	0.72
Latvia	0.31	43	0.61	0.70	0.66
Lithuania	0.40	48	0.72	0.71	0.66
Macao SAR	0.32	43	0.57	0.71	0.67
Malta	0.24	40	0.54	0.70	0.65
Morocco	0.75	66	0.79	0.79	0.86
Netherlands	0.26	40	0.10	0.78	0.77
New Zealand	0.47	50	0.62	0.76	0.73
Northern Ireland	0.36	44	0.42	0.77	0.75
Norway (5)	0.19	39	0.49	0.74	0.63
Oman	0.49	50	0.65	0.71	0.76
Poland	0.29	44	0.63	0.67	0.68
Portugal	0.44	49	0.63	0.72	0.75
Qatar	0.34	45	0.65	0.64	0.72
Russian Federation	0.55	54	0.73	0.71	0.78
Saudi Arabia	0.43	47	0.63	0.72	0.71
Singapore	0.30	44	0.66	0.67	0.66
Slovak Republic	0.45	50	0.71	0.72	0.67
Slovenia	0.34	43	0.50	0.73	0.72
South Africa	0.53	51	0.53	0.80	0.78
Spain	0.42	47	0.56	0.75	0.73
Sweden	0.21	39	0.48	0.72	0.65
Trinidad and Tobago	0.58	56	0.68	0.78	0.78
United Arab Emirates	0.26	42	0.64	0.62	0.67
United States	-	-	-	-	-
Benchmarking Participants					
Buenos Aires, Argentina	0.37	45	0.59	0.75	0.67
Ontario, Canada	0.29	42	0.56	0.69	0.69
Quebec, Canada	0.29	43	0.57	0.71	0.67
Denmark (3)	0.15	37	0.34	0.74	0.67
Norway (4)	0.22	40	0.49	0.74	0.64
Moscow City, Russian Fed.	0.24	41	0.49	0.69	0.72
Eng/Afr/Zulu - RSA (5)	0.60	55	0.67	0.78	0.78
Andalusia, Spain	0.48	50	0.58	0.76	0.76
Madrid, Spain	0.43	47	0.49	0.78	0.76
Abu Dhabi, UAE	0.24	41	0.60	0.61	0.70
Dubai, UAE	0.17	39	0.66	0.63	0.60

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

A dash (–) indicates comparable data not available.

Relationship Between the PIRLS 2016 *Digital Devices in the Home* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.12	0.02	0.01
Austria	0.12	0.01	0.02
Azerbaijan	0.29	0.09	0.06
Bahrain	0.20	0.04	0.02
Belgium (Flemish)	0.05	0.00	0.01
Belgium (French)	0.08	0.01	0.00
Bulgaria	0.36	0.13	0.09
Canada	0.12	0.01	0.01
Chile	0.23	0.05	0.04
Chinese Taipei	0.20	0.04	0.02
Czech Republic	0.14	0.02	0.03
Denmark	0.10	0.01	0.00
Egypt	0.36	0.13	0.09
England	-	-	-
Finland	0.14	0.02	0.02
France	0.03	0.00	0.00
Georgia	0.17	0.03	0.01
Germany	0.01	0.00	0.00
Hong Kong SAR	0.10	0.01	0.00
Hungary	0.26	0.07	0.04
Iran, Islamic Rep. of	0.32	0.10	0.07
Ireland	0.11	0.01	0.01
Israel	0.18	0.03	0.04
Italy	0.07	0.01	0.00
Kazakhstan	0.20	0.04	0.02
Kuwait	0.12	0.01	0.01
Latvia	0.16	0.03	0.02
Lithuania	0.22	0.05	0.04
Macao SAR	0.16	0.02	0.01
Malta	0.11	0.01	0.01
Morocco	0.31	0.10	0.07
Netherlands	0.08	0.01	0.00
New Zealand	0.18	0.03	0.02
Northern Ireland	0.12	0.01	0.01
Norway (5)	0.12	0.01	0.01
Oman	0.23	0.05	0.03
Poland	0.17	0.03	0.03
Portugal	0.21	0.04	0.03
Qatar	0.24	0.06	0.03
Russian Federation	0.28	0.08	0.04
Saudi Arabia	0.13	0.02	0.01
Singapore	0.27	0.07	0.05
Slovak Republic	0.33	0.11	0.11
Slovenia	0.18	0.03	0.02
South Africa	0.22	0.05	0.04
Spain	0.21	0.04	0.03
Sweden	0.13	0.02	0.01
Trinidad and Tobago	0.27	0.07	0.04
United Arab Emirates	0.18	0.03	0.01
United States	-	-	-
International Median	0.17	0.03	0.02
Benchmarking Participants			
Buenos Aires, Argentina	0.24	0.06	0.04
Ontario, Canada	0.11	0.01	0.01
Quebec, Canada	0.11	0.01	0.01
Denmark (3)	0.13	0.02	0.01
Norway (4)	0.16	0.03	0.02
Moscow City, Russian Fed.	0.15	0.02	0.01
Eng/Afr/Zulu - RSA (5)	0.37	0.14	0.08
Andalusia, Spain	0.18	0.03	0.02
Madrid, Spain	0.20	0.04	0.03
Abu Dhabi, UAE	0.19	0.04	0.02
Dubai, UAE	0.14	0.02	0.01

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Digital Devices in the Home* Scale and ePIRLS 2016 Online Informational Reading Achievement

Country	Pearson's Correlation with ePIRLS Achievement		Variance in ePIRLS Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Canada	0.19	0.04	0.02
Chinese Taipei	0.24	0.06	0.03
Denmark	0.10	0.01	0.00
Georgia	0.16	0.03	0.01
Ireland	0.15	0.02	0.01
Israel	0.18	0.03	0.03
Italy	0.13	0.02	0.01
Norway	0.12	0.01	0.01
Portugal	0.26	0.07	0.04
Singapore	0.29	0.08	0.06
Slovenia	0.20	0.04	0.03
Sweden	0.14	0.02	0.01
United Arab Emirates	0.18	0.03	0.01
United States	-	-	-
International Median	0.18	0.03	0.01
Benchmarking Participants			
Abu Dhabi, UAE	0.20	0.04	0.02
Dubai, UAE	0.14	0.02	0.01

A dash (–) indicates comparable data not available.

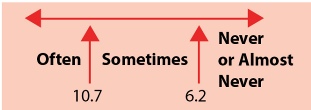
Early Literacy Activities Before Beginning Primary School Scale

The *Early Literacy Activities Before Beginning Primary School* (ELA) scale was created based on parents' frequency of doing the nine activities described below.

Items in the PIRLS 2016 *Early Literacy Activities Before Beginning Primary School* Scale

Before your child began primary/elementary school, how often did you or someone else in your home do the following activities with him or her?

			Often	Sometimes	Never or almost never
T	ASBH02A	1) Read books	◯	◯	◯
T	ASBH02B	2) Tell stories	◯	◯	◯
T	ASBH02C	3) Sing songs	◯	◯	◯
T	ASBH02D	4) Play with alphabet toys (e.g., blocks with letters of the alphabet)	◯	◯	◯
T	ASBH02E	5) Talk about things you had done	◯	◯	◯
T	ASBH02F	6) Talk about what you had read	◯	◯	◯
T	ASBH02G	7) Play word games	◯	◯	◯
T	ASBH02H	8) Write letters or words	◯	◯	◯
T	ASBH02I	9) Read aloud signs and labels	◯	◯	◯



T Trend item—item was included in the same scale in PIRLS 2011 and was used for linking the PIRLS 2011 and PIRLS 2016 scales.

Item Parameters for the PIRLS 2016 *Early Literacy Activities Before Beginning Primary School* Scale

Item	delta	tau_1	tau_2	Infit
ASBH02A	-0.43565	-1.37653	1.37653	0.96
ASBH02B	-0.21414	-1.38385	1.38385	0.97
ASBH02C	0.14115	-0.92504	0.92504	1.14
ASBH02D	0.31840	-0.99895	0.99895	1.02
ASBH02E	-0.89489	-1.18523	1.18523	1.03
ASBH02F	0.31620	-1.36183	1.36183	0.95
ASBH02G	0.54796	-1.22190	1.22190	0.95
ASBH02H	0.02688	-1.14591	1.14591	1.00
ASBH02I	0.19409	-0.98400	0.98400	1.00

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Early Literacy Activities Before Beginning Primary School* Scale

Scale Transformation Constants	
A = 8.467126	Transformed Scale Score = 8.467126 + 1.488680 • Logit Scale Score
B = 1.488680	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Early Literacy Activities Before Beginning Primary School* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	2.13736	
1	3.95698	
2	4.90763	
3	5.60873	
4	6.18976	6.2
5	6.70267	
6	7.17841	
7	7.62846	
8	8.06349	
9	8.49167	
10	8.91693	
11	9.34703	
12	9.78848	
13	10.25182	
14	10.74970	10.7
15	11.31055	
16	11.98750	
17	12.91216	
18	14.70318	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Early Literacy Activities Before Beginning Primary School* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBH02A	ASBH02B	ASBH02C	ASBH02D	ASBH02E	ASBH02F	ASBH02G	ASBH02H	ASBH02I
Australia	0.83	42	0.63	0.66	0.58	0.63	0.62	0.68	0.72	0.63	0.69
Austria	0.70	30	0.58	0.63	0.53	0.47	0.50	0.61	0.63	0.39	0.53
Azerbaijan	0.76	35	0.65	0.61	0.40	0.59	0.45	0.66	0.59	0.65	0.65
Bahrain	0.71	31	0.54	0.56	0.38	0.59	0.48	0.57	0.63	0.61	0.57
Belgium (Flemish)	0.75	33	0.55	0.56	0.51	0.56	0.51	0.60	0.67	0.59	0.62
Belgium (French)	0.72	31	0.57	0.56	0.44	0.60	0.47	0.55	0.67	0.58	0.56
Bulgaria	0.85	46	0.71	0.69	0.50	0.73	0.63	0.72	0.70	0.74	0.67
Canada	0.80	39	0.61	0.63	0.54	0.62	0.53	0.67	0.71	0.63	0.67
Chile	0.78	36	0.60	0.64	0.51	0.63	0.48	0.63	0.69	0.65	0.56
Chinese Taipei	0.81	40	0.69	0.69	0.58	0.60	0.57	0.64	0.67	0.58	0.67
Czech Republic	0.71	30	0.52	0.58	0.41	0.50	0.51	0.67	0.59	0.55	0.57
Denmark	0.75	33	0.56	0.58	0.50	0.56	0.46	0.67	0.66	0.59	0.58
Egypt	0.85	46	0.66	0.70	0.51	0.73	0.63	0.71	0.72	0.73	0.67
England	-	-	-	-	-	-	-	-	-	-	-
Finland	0.74	33	0.56	0.58	0.48	0.58	0.46	0.68	0.60	0.59	0.59
France	0.71	30	0.55	0.55	0.41	0.58	0.46	0.58	0.63	0.59	0.56
Georgia	0.77	37	0.66	0.59	0.42	0.67	0.46	0.61	0.65	0.64	0.69
Germany	0.69	30	0.54	0.57	0.46	0.48	0.47	0.61	0.66	0.51	0.57
Hong Kong SAR	0.78	36	0.63	0.65	0.59	0.49	0.55	0.66	0.69	0.53	0.62
Hungary	0.73	31	0.51	0.49	0.48	0.66	0.45	0.55	0.64	0.60	0.64
Iran, Islamic Rep. of	0.79	37	0.62	0.54	0.50	0.66	0.54	0.63	0.67	0.66	0.65
Ireland	0.80	38	0.58	0.63	0.54	0.59	0.52	0.65	0.68	0.66	0.68
Israel	0.76	35	0.53	0.58	0.47	0.65	0.49	0.64	0.69	0.63	0.60
Italy	0.71	30	0.54	0.57	0.41	0.61	0.41	0.59	0.61	0.58	0.58
Kazakhstan	0.70	30	0.60	0.44	0.40	0.54	0.54	0.63	0.48	0.60	0.61
Kuwait	0.72	31	0.54	0.51	0.35	0.62	0.49	0.55	0.66	0.64	0.59
Latvia	0.72	31	0.62	0.59	0.44	0.57	0.47	0.64	0.62	0.55	0.51
Lithuania	0.72	32	0.57	0.54	0.44	0.64	0.52	0.68	0.63	0.57	0.46
Macao SAR	0.78	36	0.65	0.65	0.57	0.52	0.53	0.64	0.65	0.58	0.61
Malta	0.80	38	0.66	0.65	0.52	0.59	0.53	0.69	0.67	0.60	0.64
Morocco	0.86	46	0.71	0.66	0.43	0.71	0.62	0.73	0.71	0.80	0.70
Netherlands	0.73	32	0.48	0.52	0.48	0.57	0.47	0.55	0.69	0.65	0.61
New Zealand	0.83	42	0.64	0.67	0.57	0.64	0.57	0.69	0.73	0.65	0.69
Northern Ireland	0.81	40	0.57	0.64	0.53	0.61	0.58	0.71	0.72	0.62	0.65
Norway (5)	0.76	34	0.57	0.59	0.46	0.62	0.43	0.65	0.67	0.59	0.62
Oman	0.70	31	0.59	0.55	0.24	0.57	0.46	0.58	0.62	0.62	0.63
Poland	0.76	34	0.55	0.57	0.50	0.62	0.54	0.67	0.61	0.60	0.58
Portugal	0.76	34	0.63	0.65	0.48	0.62	0.49	0.62	0.67	0.58	0.49
Qatar	0.77	35	0.60	0.59	0.49	0.60	0.52	0.62	0.64	0.64	0.63
Russian Federation	0.77	35	0.61	0.53	0.49	0.64	0.58	0.66	0.64	0.62	0.56
Saudi Arabia	0.72	32	0.59	0.60	0.22	0.60	0.51	0.61	0.65	0.65	0.57
Singapore	0.83	43	0.70	0.72	0.58	0.65	0.58	0.68	0.67	0.64	0.69
Slovak Republic	0.76	35	0.58	0.57	0.43	0.62	0.62	0.70	0.62	0.57	0.56
Slovenia	0.75	34	0.53	0.62	0.55	0.59	0.44	0.60	0.66	0.55	0.64
South Africa	0.77	36	0.64	0.51	0.46	0.64	0.52	0.64	0.65	0.66	0.66
Spain	0.75	34	0.59	0.60	0.46	0.63	0.45	0.60	0.66	0.62	0.56
Sweden	0.76	35	0.56	0.54	0.49	0.61	0.51	0.67	0.68	0.62	0.60
Trinidad and Tobago	0.78	36	0.67	0.65	0.51	0.55	0.50	0.66	0.62	0.61	0.63
United Arab Emirates	0.73	32	0.54	0.56	0.44	0.55	0.54	0.58	0.63	0.61	0.59
United States	-	-	-	-	-	-	-	-	-	-	-
Benchmarking Participants											
Buenos Aires, Argentina	0.76	34	0.62	0.62	0.49	0.66	0.46	0.58	0.67	0.62	0.52
Ontario, Canada	0.81	40	0.62	0.64	0.56	0.61	0.52	0.68	0.72	0.63	0.70
Quebec, Canada	0.76	35	0.58	0.60	0.48	0.62	0.51	0.61	0.68	0.59	0.62
Denmark (3)	0.74	33	0.56	0.59	0.50	0.55	0.47	0.67	0.65	0.55	0.56
Norway (4)	0.76	34	0.55	0.59	0.48	0.63	0.45	0.63	0.66	0.61	0.61
Moscow City, Russian Fed.	0.76	35	0.59	0.58	0.53	0.59	0.59	0.67	0.64	0.58	0.50
Eng/Afr/Zulu - RSA (5)	0.75	34	0.63	0.58	0.42	0.60	0.49	0.58	0.63	0.64	0.62
Andalusia, Spain	0.75	34	0.59	0.60	0.49	0.62	0.47	0.61	0.64	0.61	0.56
Madrid, Spain	0.73	32	0.56	0.58	0.47	0.57	0.42	0.62	0.63	0.61	0.57
Abu Dhabi, UAE	0.73	32	0.53	0.55	0.44	0.54	0.57	0.59	0.63	0.63	0.60
Dubai, UAE	0.75	33	0.60	0.59	0.45	0.57	0.52	0.61	0.63	0.60	0.61

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Early Literacy Activities Before Beginning Primary School* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.15	0.02	0.02
Austria	0.20	0.04	0.03
Azerbaijan	0.16	0.02	0.02
Bahrain	0.22	0.05	0.04
Belgium (Flemish)	0.18	0.03	0.03
Belgium (French)	0.19	0.03	0.03
Bulgaria	0.41	0.17	0.15
Canada	0.19	0.04	0.03
Chile	0.24	0.06	0.04
Chinese Taipei	0.24	0.06	0.04
Czech Republic	0.13	0.02	0.01
Denmark	0.19	0.04	0.03
Egypt	0.34	0.12	0.09
England	-	-	-
Finland	0.18	0.03	0.02
France	0.17	0.03	0.02
Georgia	0.08	0.01	0.01
Germany	0.18	0.03	0.02
Hong Kong SAR	0.08	0.01	0.00
Hungary	0.14	0.02	0.02
Iran, Islamic Rep. of	0.24	0.06	0.05
Ireland	0.27	0.07	0.06
Israel	0.16	0.02	0.02
Italy	0.15	0.02	0.01
Kazakhstan	0.10	0.01	0.00
Kuwait	0.19	0.04	0.02
Latvia	0.13	0.02	0.01
Lithuania	0.14	0.02	0.02
Macao SAR	0.10	0.01	0.01
Malta	0.19	0.04	0.03
Morocco	0.21	0.04	0.04
Netherlands	0.14	0.02	0.01
New Zealand	0.26	0.07	0.06
Northern Ireland	0.16	0.03	0.03
Norway (5)	0.18	0.03	0.03
Oman	0.25	0.06	0.04
Poland	0.12	0.01	0.01
Portugal	0.19	0.04	0.03
Qatar	0.22	0.05	0.04
Russian Federation	0.17	0.03	0.02
Saudi Arabia	0.16	0.03	0.02
Singapore	0.24	0.06	0.04
Slovak Republic	0.25	0.06	0.08
Slovenia	0.21	0.04	0.04
South Africa	0.11	0.01	0.02
Spain	0.23	0.05	0.04
Sweden	0.18	0.03	0.03
Trinidad and Tobago	0.25	0.06	0.05
United Arab Emirates	0.25	0.06	0.04
United States	-	-	-
International Median	0.19	0.03	0.03
Benchmarking Participants			
Buenos Aires, Argentina	0.26	0.07	0.05
Ontario, Canada	0.20	0.04	0.03
Quebec, Canada	0.17	0.03	0.03
Denmark (3)	0.16	0.02	0.02
Norway (4)	0.20	0.04	0.03
Moscow City, Russian Fed.	0.17	0.03	0.02
Eng/Afr/Zulu - RSA (5)	0.13	0.02	0.02
Andalusia, Spain	0.24	0.06	0.04
Madrid, Spain	0.17	0.03	0.02
Abu Dhabi, UAE	0.23	0.05	0.04
Dubai, UAE	0.27	0.07	0.05

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Home Resources for Learning Scale

The *Home Resources for Learning* (HRL) scale was created based on students' and parents' responses concerning the availability of five resources described below.

Items in the PIRLS 2016 *Home Resources for Learning* Scale

T | ASBG04

Number of books in the home (students):
1) 0-10
2) 11-25
3) 26-100
4) 101-200
5) More than 200

T | ASBH14

Number of children's books in the home (parents):
1) 0-10
2) 11-25
3) 26-50
4) 51-100
5) More than 100

T | ASDG05S[1]

Number of home study supports (students):
1) None
2) Internet connection or own room
3) Both

T | ASDHEDUP[1]

Highest level of education of either parent (parents):
1) Finished some primary or lower secondary or did not go to school
2) Finished lower secondary
3) Finished upper secondary
4) Finished post-secondary education
5) Finished university or higher

T | ASDHOCCP[1]

Highest level of occupation of either parent (parents):
1) Has never worked outside home for pay, general laborer, or semi-professional (skilled agricultural or fishery worker, craft or trade worker, plant or machine operator)
2) Clerical (clerk or service or sales worker)
3) Small business owner
4) Professional (corporate manager or senior official, professional, or technician or associate professional)

Many Resources | **Some Resources** | **Few Resources**

11.8 | 7.5

T Trend item—item was included in the same scale in PIRLS 2011 and was used for linking the PIRLS 2011 and PIRLS 2016 scales.
1 Derived variable. For more details, see Supplement 3 of the *User Guide for the PIRLS 2016 International Database*.

Item Parameters for the PIRLS 2016 *Home Resources for Learning* Scale

Item	delta	tau_1	tau_2	tau_3	tau_4	Infit
ASBG04	0.64203	-1.31056	-0.54699	0.96645	0.89110	1.00
ASDG05S	-0.95160	-0.80498	0.80498			1.06
ASDHOCCP	0.08290	-0.31922	0.95822	-0.63900		1.02
ASBH14	0.66949	-0.85830	-0.54569	0.43193	0.97206	0.96
ASDHEDUP	-0.44282	-0.60843	-0.91684	0.99045	0.53482	0.96

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Home Resources for Learning* Scale

Scale Transformation Constants	
A = 9.470169	Transformed Scale Score = 9.470169 + 1.658175 • Logit Scale Score
B = 1.658175	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Home Resources for Learning* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	4.03041	
1	5.88662	
2	6.79193	
3	7.44990	7.5
4	7.99030	
5	8.45482	
6	8.86479	
7	9.23157	
8	9.56865	
9	9.89060	
10	10.22039	
11	10.55954	
12	10.92762	
13	11.34044	
14	11.82688	11.8
15	12.41852	
16	13.22250	
17	14.80246	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Home Resources for Learning* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			ASBG04	ASDG05S	ASDHOCCP	ASBH14	ASDHEDUP
Australia	0.65	42	0.66	0.31	0.76	0.69	0.73
Austria	0.73	48	0.75	0.31	0.75	0.77	0.78
Azerbaijan	0.65	42	0.62	0.54	0.71	0.61	0.73
Bahrain	0.59	38	0.54	0.22	0.73	0.67	0.77
Belgium (Flemish)	0.67	44	0.70	0.31	0.72	0.71	0.76
Belgium (French)	0.72	48	0.71	0.35	0.75	0.78	0.77
Bulgaria	0.82	58	0.77	0.44	0.82	0.82	0.87
Canada	0.61	40	0.68	0.31	0.70	0.69	0.67
Chile	0.69	45	0.57	0.39	0.79	0.70	0.81
Chinese Taipei	0.73	48	0.78	0.27	0.72	0.78	0.78
Czech Republic	0.68	45	0.73	0.12	0.74	0.76	0.76
Denmark	0.66	44	0.70	0.31	0.74	0.72	0.74
Egypt	0.62	40	0.53	0.55	0.72	0.54	0.77
England	-	-	-	-	-	-	-
Finland	0.64	42	0.69	0.18	0.72	0.70	0.75
France	0.72	47	0.73	0.42	0.70	0.77	0.76
Georgia	0.71	47	0.71	0.45	0.71	0.77	0.75
Germany	0.70	46	0.70	0.28	0.71	0.79	0.78
Hong Kong SAR	0.74	49	0.73	0.36	0.77	0.77	0.79
Hungary	0.81	58	0.81	0.31	0.83	0.83	0.87
Iran, Islamic Rep. of	0.73	48	0.70	0.59	0.67	0.71	0.78
Ireland	0.68	45	0.72	0.11	0.74	0.75	0.77
Israel	0.73	56	0.63	-	0.78	0.76	0.81
Italy	0.67	44	0.67	0.13	0.76	0.75	0.79
Kazakhstan	0.60	39	0.64	0.53	0.65	0.63	0.67
Kuwait	0.53	35	0.47	0.23	0.74	0.55	0.79
Latvia	0.68	45	0.69	0.29	0.70	0.77	0.77
Lithuania	0.74	50	0.72	0.36	0.78	0.79	0.77
Macao SAR	0.71	47	0.72	0.29	0.74	0.75	0.78
Malta	0.62	41	0.55	0.09	0.80	0.66	0.81
Morocco	0.76	50	0.71	0.60	0.75	0.68	0.78
Netherlands	0.67	44	0.68	0.16	0.78	0.69	0.79
New Zealand	0.67	44	0.67	0.36	0.72	0.72	0.76
Northern Ireland	0.68	45	0.71	0.23	0.75	0.71	0.79
Norway (5)	0.67	45	0.67	0.28	0.75	0.73	0.79
Oman	0.63	41	0.53	0.42	0.75	0.62	0.78
Poland	0.73	48	0.68	0.33	0.79	0.73	0.82
Portugal	0.73	48	0.70	0.20	0.79	0.77	0.81
Qatar	0.56	37	0.53	0.32	0.68	0.68	0.72
Russian Federation	0.64	42	0.66	0.39	0.69	0.69	0.75
Saudi Arabia	0.51	34	0.49	0.22	0.73	0.51	0.77
Singapore	0.67	44	0.70	0.29	0.72	0.72	0.77
Slovak Republic	0.77	53	0.77	0.36	0.80	0.80	0.81
Slovenia	0.69	46	0.70	0.07	0.77	0.75	0.80
South Africa	0.62	39	0.56	0.55	0.74	0.50	0.71
Spain	0.68	45	0.66	0.14	0.78	0.75	0.78
Sweden	0.69	46	0.71	0.37	0.73	0.75	0.76
Trinidad and Tobago	0.61	39	0.60	0.42	0.71	0.62	0.72
United Arab Emirates	0.58	38	0.56	0.35	0.68	0.71	0.69
United States	-	-	-	-	-	-	-
Benchmarking Participants							
Buenos Aires, Argentina	0.74	49	0.62	0.35	0.82	0.77	0.81
Ontario, Canada	0.57	37	0.64	0.33	0.68	0.67	0.66
Quebec, Canada	0.63	41	0.70	0.31	0.70	0.72	0.66
Denmark (3)	0.65	43	0.69	0.24	0.74	0.72	0.76
Norway (4)	0.66	44	0.64	0.25	0.74	0.72	0.79
Moscow City, Russian Fed.	0.59	39	0.63	0.24	0.70	0.73	0.70
Eng/Afr/Zulu - RSA (5)	0.70	45	0.64	0.56	0.75	0.60	0.76
Andalusia, Spain	0.69	46	0.67	0.11	0.77	0.75	0.81
Madrid, Spain	0.69	46	0.68	0.09	0.80	0.73	0.80
Abu Dhabi, UAE	0.54	36	0.51	0.31	0.69	0.68	0.71
Dubai, UAE	0.60	39	0.67	0.40	0.61	0.76	0.63

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Home Resources for Learning* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.41	0.17	0.12
Austria	0.49	0.24	0.13
Azerbaijan	0.31	0.10	0.06
Bahrain	0.37	0.13	0.06
Belgium (Flemish)	0.39	0.15	0.10
Belgium (French)	0.47	0.22	0.16
Bulgaria	0.56	0.31	0.23
Canada	0.35	0.12	0.09
Chile	0.34	0.11	0.05
Chinese Taipei	0.39	0.15	0.09
Czech Republic	0.44	0.20	0.14
Denmark	0.37	0.14	0.10
Egypt	0.40	0.16	0.13
England	-	-	-
Finland	0.38	0.14	0.09
France	0.42	0.18	0.11
Georgia	0.33	0.11	0.06
Germany	0.47	0.22	0.12
Hong Kong SAR	0.11	0.01	0.01
Hungary	0.55	0.31	0.22
Iran, Islamic Rep. of	0.45	0.20	0.14
Ireland	0.44	0.19	0.14
Israel	0.47	0.22	0.12
Italy	0.36	0.13	0.07
Kazakhstan	0.25	0.06	0.03
Kuwait	0.25	0.06	0.02
Latvia	0.37	0.14	0.07
Lithuania	0.45	0.20	0.11
Macao SAR	0.22	0.05	0.04
Malta	0.23	0.05	0.03
Morocco	0.34	0.12	0.08
Netherlands	0.36	0.13	0.08
New Zealand	0.42	0.18	0.13
Northern Ireland	0.38	0.14	0.10
Norway (5)	0.35	0.13	0.08
Oman	0.36	0.13	0.07
Poland	0.41	0.17	0.10
Portugal	0.37	0.14	0.09
Qatar	0.39	0.15	0.07
Russian Federation	0.39	0.15	0.07
Saudi Arabia	0.15	0.02	0.01
Singapore	0.51	0.26	0.16
Slovak Republic	0.59	0.35	0.30
Slovenia	0.44	0.19	0.11
South Africa	0.38	0.14	0.09
Spain	0.36	0.13	0.09
Sweden	0.41	0.17	0.12
Trinidad and Tobago	0.36	0.13	0.07
United Arab Emirates	0.42	0.18	0.09
United States	-	-	-
International Median	0.38	0.15	0.09
Benchmarking Participants			
Buenos Aires, Argentina	0.42	0.18	0.10
Ontario, Canada	0.33	0.11	0.09
Quebec, Canada	0.34	0.11	0.08
Denmark (3)	0.30	0.09	0.06
Norway (4)	0.35	0.13	0.10
Moscow City, Russian Fed.	0.33	0.11	0.07
Eng/Afr/Zulu - RSA (5)	0.46	0.21	0.13
Andalusia, Spain	0.40	0.16	0.09
Madrid, Spain	0.35	0.12	0.08
Abu Dhabi, UAE	0.40	0.16	0.08
Dubai, UAE	0.45	0.20	0.12

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

A dash (–) indicates comparable data not available.

Relationship Between the PIRLS 2016 *Home Resources for Learning* Scale and ePIRLS 2016 Online Informational Reading Achievement

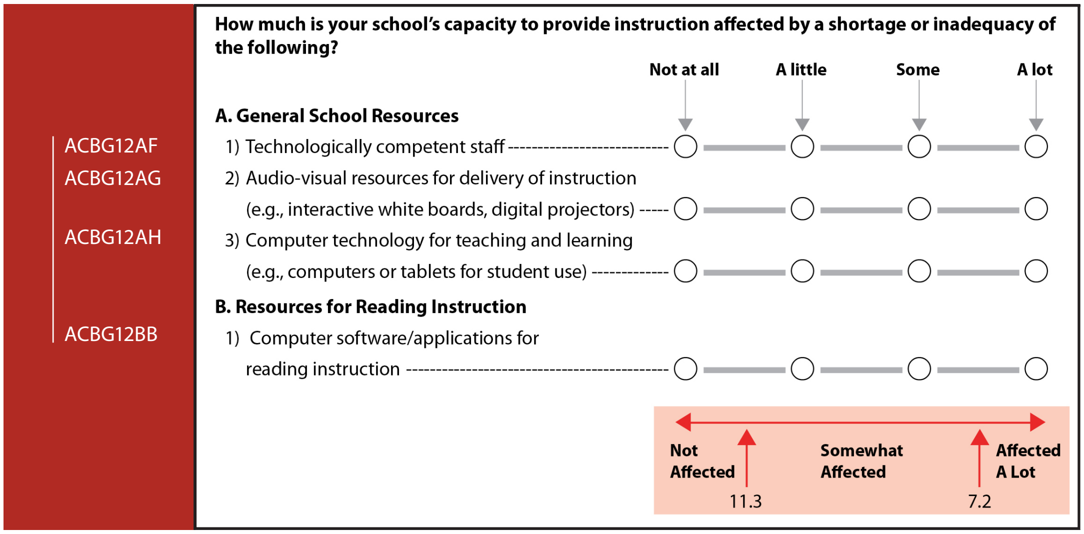
Country	Pearson's Correlation with ePIRLS Achievement		Variance in ePIRLS Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Canada	0.33	0.11	0.09
Chinese Taipei	0.35	0.13	0.08
Denmark	0.34	0.12	0.10
Georgia	0.27	0.07	0.04
Ireland	0.43	0.18	0.13
Israel	0.44	0.19	0.10
Italy	0.32	0.10	0.06
Norway	0.33	0.11	0.07
Portugal	0.37	0.14	0.09
Singapore	0.49	0.24	0.15
Slovenia	0.39	0.15	0.09
Sweden	0.41	0.17	0.12
United Arab Emirates	0.40	0.16	0.08
United States	-	-	-
International Median	0.37	0.14	0.09
Benchmarking Participants			
Abu Dhabi, UAE	0.39	0.15	0.07
Dubai, UAE	0.44	0.19	0.11

A dash (–) indicates comparable data not available.

Instruction Affected by Digital Resource Shortages Scale

The *Instruction Affected by Digital Resource Shortages* (DRS) scale was created based on principals' responses concerning four school and classroom resources described below.

Items in the PIRLS 2016 *Instruction Affected by Digital Resource Shortages* Scale



How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following?

		Not at all	A little	Some	A lot
	A. General School Resources				
ACBG12AF	1) Technologically competent staff	◯	◯	◯	◯
ACBG12AG	2) Audio-visual resources for delivery of instruction (e.g., interactive white boards, digital projectors)	◯	◯	◯	◯
ACBG12AH	3) Computer technology for teaching and learning (e.g., computers or tablets for student use)	◯	◯	◯	◯
	B. Resources for Reading Instruction				
ACBG12BB	1) Computer software/applications for reading instruction	◯	◯	◯	◯

Item Parameters for the PIRLS 2016 *Instruction Affected by Digital Resource Shortages* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ACBG12AF	-0.06232	-1.40083	-0.09454	1.49537	1.07
ACBG12AG	-0.14345	-1.27363	0.17090	1.10273	0.87
ACBG12AH	0.16006	-1.46129	0.04607	1.41522	0.88
ACBG12BB	0.04571	-1.73178	-0.08308	1.81486	1.14

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Instruction Affected by Digital Resource Shortages* Scale

Scale Transformation Constants	
A = 9.264589	Transformed Scale Score = 9.264589 + 1.227112 • Logit Scale Score
B = 1.227112	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Instruction Affected by Digital Resource Shortages* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	4.82238	
1	6.34762	
2	7.16657	7.2
3	7.78270	
4	8.31408	
5	8.79730	
6	9.25885	
7	9.72012	
8	10.20216	
9	10.73577	
10	11.36104	11.3
11	12.19853	
12	13.75008	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Instruction Affected by Digital Resource Shortages* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item: ACBG12AF	ACBG12AG	ACBG12AH	ACBG12BB
Australia	0.85	69	0.82	0.82	0.88	0.79
Austria	0.74	57	0.60	0.81	0.85	0.74
Azerbaijan	0.76	59	0.70	0.78	0.82	0.76
Bahrain	0.87	73	0.84	0.89	0.88	0.79
Belgium (Flemish)	0.68	51	0.71	0.75	0.81	0.56
Belgium (French)	0.73	56	0.82	0.78	0.73	0.65
Bulgaria	0.69	52	0.50	0.81	0.86	0.65
Canada	0.85	68	0.82	0.85	0.88	0.76
Chile	0.72	56	0.65	0.84	0.88	0.59
Chinese Taipei	0.83	67	0.84	0.91	0.90	0.57
Czech Republic	0.69	52	0.56	0.76	0.84	0.70
Denmark	0.70	53	0.61	0.81	0.72	0.77
Egypt	0.78	61	0.66	0.82	0.81	0.82
England	0.79	62	0.76	0.84	0.84	0.70
Finland	0.76	58	0.72	0.78	0.82	0.71
France	0.72	56	0.52	0.82	0.89	0.72
Georgia	0.72	53	0.63	0.77	0.75	0.76
Germany	0.78	61	0.71	0.81	0.83	0.76
Hong Kong SAR	0.70	52	0.74	0.86	0.76	0.48
Hungary	0.67	52	0.42	0.87	0.83	0.69
Iran, Islamic Rep. of	0.77	60	0.70	0.86	0.85	0.66
Ireland	0.71	54	0.72	0.64	0.81	0.75
Israel	0.85	68	0.82	0.85	0.83	0.80
Italy	0.77	59	0.82	0.83	0.82	0.59
Kazakhstan	0.83	68	0.73	0.85	0.85	0.86
Kuwait	0.77	59	0.74	0.86	0.75	0.72
Latvia	0.75	59	0.82	0.88	0.85	0.47
Lithuania	0.77	59	0.68	0.84	0.83	0.72
Macao SAR	0.70	58	0.86	0.88	0.91	-0.07
Malta	0.80	62	0.75	0.85	0.84	0.71
Morocco	0.86	71	0.75	0.87	0.89	0.85
Netherlands	0.58	45	0.52	0.69	0.76	0.70
New Zealand	0.84	67	0.79	0.84	0.89	0.75
Northern Ireland	0.81	64	0.81	0.86	0.80	0.72
Norway (5)	0.74	56	0.69	0.79	0.79	0.73
Oman	0.81	64	0.80	0.84	0.84	0.72
Poland	0.72	55	0.69	0.83	0.86	0.56
Portugal	0.81	63	0.76	0.76	0.85	0.81
Qatar	0.94	84	0.93	0.96	0.94	0.85
Russian Federation	0.81	64	0.76	0.86	0.85	0.73
Saudi Arabia	0.78	60	0.74	0.81	0.81	0.75
Singapore	0.90	77	0.87	0.91	0.94	0.79
Slovak Republic	0.74	56	0.65	0.85	0.81	0.66
Slovenia	0.69	53	0.60	0.86	0.84	0.57
South Africa	0.88	73	0.66	0.91	0.92	0.90
Spain	0.80	63	0.78	0.84	0.86	0.69
Sweden	0.78	61	0.78	0.81	0.77	0.76
Trinidad and Tobago	0.69	52	0.41	0.70	0.87	0.83
United Arab Emirates	0.89	75	0.88	0.91	0.90	0.75
United States	0.83	67	0.84	0.82	0.85	0.75
Benchmarking Participants						
Buenos Aires, Argentina	0.82	65	0.77	0.82	0.83	0.79
Ontario, Canada	0.87	72	0.84	0.89	0.88	0.79
Quebec, Canada	0.79	61	0.78	0.74	0.87	0.73
Denmark (3)	0.70	53	0.62	0.81	0.73	0.73
Norway (4)	0.73	55	0.69	0.77	0.79	0.73
Moscow City, Russian Fed.	0.86	71	0.85	0.87	0.91	0.73
Eng/Afr/Zulu - RSA (5)	0.84	68	0.59	0.90	0.89	0.89
Andalusia, Spain	0.70	53	0.68	0.79	0.75	0.68
Madrid, Spain	0.80	63	0.76	0.80	0.87	0.76
Abu Dhabi, UAE	0.86	71	0.86	0.89	0.91	0.70
Dubai, UAE	0.92	81	0.90	0.93	0.92	0.85

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Instruction Affected by Digital Resource Shortages* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.11	0.01	0.01
Austria	-0.01	0.00	0.00
Azerbaijan	-0.17	0.03	0.03
Bahrain	0.11	0.01	0.01
Belgium (Flemish)	-0.01	0.00	0.00
Belgium (French)	0.06	0.00	0.00
Bulgaria	0.13	0.02	0.02
Canada	0.02	0.00	0.00
Chile	0.11	0.01	0.02
Chinese Taipei	0.09	0.01	0.00
Czech Republic	-0.04	0.00	0.01
Denmark	0.05	0.00	0.00
Egypt	0.10	0.01	0.02
England	0.05	0.00	0.00
Finland	0.04	0.00	0.01
France	0.04	0.00	0.00
Georgia	0.12	0.01	0.01
Germany	0.12	0.01	0.01
Hong Kong SAR	0.11	0.01	0.01
Hungary	-0.05	0.00	0.01
Iran, Islamic Rep. of	0.13	0.02	0.02
Ireland	0.08	0.01	0.01
Israel	0.24	0.06	0.04
Italy	-0.03	0.00	0.00
Kazakhstan	0.04	0.00	0.01
Kuwait	0.02	0.00	0.01
Latvia	-0.01	0.00	0.00
Lithuania	0.05	0.00	0.00
Macao SAR	-0.13	0.02	0.02
Malta	0.05	0.00	0.00
Morocco	-0.07	0.01	0.01
Netherlands	0.06	0.00	0.00
New Zealand	0.10	0.01	0.00
Northern Ireland	-0.01	0.00	0.00
Norway (5)	0.06	0.00	0.00
Oman	0.08	0.01	0.01
Poland	0.10	0.01	0.01
Portugal	0.11	0.01	0.01
Qatar	0.24	0.06	0.05
Russian Federation	0.08	0.01	0.00
Saudi Arabia	-0.11	0.01	0.02
Singapore	-0.08	0.01	0.01
Slovak Republic	0.04	0.00	0.02
Slovenia	0.04	0.00	0.00
South Africa	-0.02	0.00	0.03
Spain	0.16	0.03	0.03
Sweden	0.12	0.01	0.02
Trinidad and Tobago	0.09	0.01	0.01
United Arab Emirates	0.26	0.07	0.07
United States	-0.06	0.00	0.01
International Median	0.06	0.01	0.01
Benchmarking Participants			
Buenos Aires, Argentina	0.10	0.01	0.01
Ontario, Canada	-0.01	0.00	0.00
Quebec, Canada	0.02	0.00	0.00
Denmark (3)	0.02	0.00	0.00
Norway (4)	0.09	0.01	0.00
Moscow City, Russian Fed.	0.01	0.00	0.00
Eng/Afr/Zulu - RSA (5)	-0.03	0.00	0.00
Andalusia, Spain	0.13	0.02	0.01
Madrid, Spain	0.07	0.00	0.00
Abu Dhabi, UAE	0.28	0.08	0.06
Dubai, UAE	0.19	0.04	0.06

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Instruction Affected by Digital Resource Shortages* Scale and ePIRLS 2016 Online Informational Reading Achievement

Country	Pearson's Correlation with ePIRLS Achievement		Variance in ePIRLS Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Canada	0.05	0.00	0.00
Chinese Taipei	0.06	0.00	0.00
Denmark	0.02	0.00	0.00
Georgia	0.08	0.01	0.00
Ireland	0.10	0.01	0.01
Israel	0.25	0.06	0.04
Italy	0.02	0.00	0.00
Norway	0.06	0.00	0.00
Portugal	0.12	0.01	0.02
Singapore	-0.07	0.00	0.01
Slovenia	0.04	0.00	0.00
Sweden	0.14	0.02	0.02
United Arab Emirates	0.22	0.05	0.07
United States	-0.03	0.00	0.01
International Median	0.06	0.00	0.01
Benchmarking Participants			
Abu Dhabi, UAE	0.28	0.08	0.08
Dubai, UAE	0.09	0.01	0.03

Instruction Affected by Reading Resource Shortages Scale

The *Instruction Affected by Reading Resource Shortages* (RRS) scale was created based on principals' responses concerning 12 school and classroom resources described below.

Items in the PIRLS 2016 *Instruction Affected by Reading Resource Shortages* Scale

		How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following?	Not at all	A little	Some	A lot
		A. General School Resources				
T	ACBG12AA	1) Instructional materials (e.g., textbooks)	◯	◯	◯	◯
T	ACBG12AB	2) Supplies (e.g., papers, pencils, materials)	◯	◯	◯	◯
T	ACBG12AC	3) School buildings and grounds	◯	◯	◯	◯
T	ACBG12AD	4) Heating/cooling and lighting systems	◯	◯	◯	◯
T	ACBG12AE	5) Instructional space (e.g., classrooms)	◯	◯	◯	◯
T	ACBG12AF	6) Technologically competent staff	◯	◯	◯	◯
	ACBG12AG	7) Audio-visual resources for delivery of instruction (e.g., interactive white boards, digital projectors)	◯	◯	◯	◯
	ACBG12AH	8) Computer technology for teaching and learning (e.g., computers or tablets for student use)	◯	◯	◯	◯
		B. Resources for Reading Instruction				
T	ACBG12BA	1) Teachers with a specialization in reading	◯	◯	◯	◯
T	ACBG12BB	2) Computer software/applications for reading instruction	◯	◯	◯	◯
	ACBG12BC	3) Library resources (books, ebooks, magazines, etc.)	◯	◯	◯	◯
	ACBG12BD	4) Instructional materials for reading (e.g., reading series, textbooks)	◯	◯	◯	◯

Not Affected — 10.8 — **Somewhat Affected** — 7.1 — **Affected A Lot**

T Trend item—item was included in the same scale in PIRLS 2011 and was used for linking the PIRLS 2011 and PIRLS 2016 scales.

Item Parameters for the PIRLS 2016 *Instruction Affected by Reading Resource Shortages* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ACBG12AA	-0.24046	-0.50371	0.08908	0.41463	0.82
ACBG12AB	-0.61342	-0.65578	0.29517	0.36061	0.80
ACBG12AC	0.08419	-0.83453	0.08039	0.75414	0.99
ACBG12AD	-0.24529	-0.51694	0.09957	0.41737	0.89
ACBG12AE	0.10860	-0.60282	-0.03207	0.63489	0.96
ACBG12AF	0.21641	-1.33100	-0.06590	1.39690	0.93
ACBG12AG	0.13549	-1.21210	0.19877	1.01333	1.01
ACBG12AH	0.42199	-1.38470	0.07125	1.31345	1.11
ACBG12BA	-0.06206	-0.90262	-0.08571	0.98833	1.16
ACBG12BB	0.31635	-1.64938	-0.05748	1.70686	1.28
ACBG12BC	0.03715	-1.42224	0.02144	1.40080	0.92
ACBG12BD	-0.15895	-1.09436	0.07055	1.02381	0.75

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Instruction Affected by Reading Resource Shortages* Scale

Scale Transformation Constants	
A = 8.945066	Transformed Scale Score = 8.945066 + 1.274387 • Logit Scale Score
B = 1.274387	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Instruction Affected by Reading Resource Shortages* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.54611	
1	4.98767	
2	5.67829	
3	6.14581	
4	6.50349	
5	6.79540	
6	7.04460	7.1
7	7.26319	
8	7.45917	
9	7.63803	
10	7.80296	
11	7.95892	
12	8.10685	
13	8.24846	
14	8.38527	
15	8.51855	
16	8.64954	
17	8.77923	
18	8.90734	
19	9.03876	
20	9.17049	
21	9.30484	
22	9.44284	
23	9.58565	
24	9.73450	
25	9.89005	
26	10.05541	
27	10.23239	
28	10.42394	
29	10.63402	
30	10.86810	10.8
31	11.13383	
32	11.44471	
33	11.82273	
34	12.31260	
35	13.02750	
36	14.49613	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Instruction Affected by Reading Resource Shortages* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			ACBG12AA	ACBG12AB	ACBG12AC	ACBG12AD	ACBG12AE	ACBG12AF	ACBG12AG	ACBG12AH	ACBG12BA	ACBG12BB	ACBG12BC	ACBG12BD
Australia	0.93	57	0.76	0.77	0.76	0.69	0.69	0.73	0.78	0.75	0.71	0.75	0.81	0.84
Austria	0.80	33	0.45	0.44	0.64	0.33	0.62	0.48	0.66	0.69	0.32	0.69	0.66	0.67
Azerbaijan	0.91	50	0.71	0.71	0.72	0.77	0.75	0.70	0.56	0.68	0.69	0.64	0.74	0.77
Bahrain	0.96	70	0.89	0.88	0.83	0.89	0.91	0.81	0.86	0.77	0.76	0.67	0.88	0.89
Belgium (Flemish)	0.81	34	0.66	0.55	0.53	0.55	0.57	0.59	0.63	0.63	0.49	0.53	0.60	0.66
Belgium (French)	0.79	31	0.60	0.57	0.49	0.63	0.46	0.63	0.63	0.50	0.52	0.61	0.51	0.55
Bulgaria	0.82	43	0.85	0.74	0.68	0.88	0.77	0.44	0.39	0.35	0.83	-0.04	0.43	0.87
Canada	0.90	49	0.71	0.73	0.64	0.58	0.61	0.71	0.76	0.73	0.62	0.69	0.78	0.78
Chile	0.91	51	0.77	0.80	0.69	0.53	0.74	0.69	0.70	0.76	0.62	0.55	0.81	0.83
Chinese Taipei	0.94	60	0.78	0.73	0.80	0.88	0.82	0.85	0.85	0.79	0.59	0.47	0.79	0.85
Czech Republic	0.78	31	0.72	0.54	0.52	0.56	0.54	0.51	0.52	0.61	0.57	0.50	0.33	0.70
Denmark	0.85	40	0.66	0.58	0.66	0.59	0.52	0.51	0.66	0.54	0.59	0.69	0.68	0.79
Egypt	0.87	41	0.57	0.72	0.73	0.61	0.69	0.56	0.53	0.52	0.67	0.62	0.66	0.75
England	0.87	43	0.69	0.68	0.68	0.54	0.59	0.61	0.80	0.75	0.38	0.58	0.65	0.78
Finland	0.88	44	0.77	0.72	0.66	0.66	0.73	0.64	0.65	0.64	0.58	0.61	0.53	0.69
France	0.79	31	0.68	0.64	0.58	0.57	0.49	0.42	0.44	0.52	0.43	0.56	0.62	0.69
Georgia	0.88	45	0.69	0.78	0.68	0.72	0.70	0.59	0.51	0.76	0.61	0.57	0.66	0.69
Germany	0.88	43	0.70	0.70	0.62	0.45	0.61	0.67	0.65	0.65	0.61	0.69	0.71	0.73
Hong Kong SAR	0.85	38	0.54	0.51	0.74	0.67	0.80	0.72	0.67	0.58	0.53	0.38	0.55	0.62
Hungary	0.88	44	0.76	0.71	0.70	0.73	0.73	0.47	0.71	0.56	0.58	0.49	0.68	0.74
Iran, Islamic Rep. of	0.89	46	0.70	0.80	0.79	0.85	0.85	0.69	0.58	0.43	0.75	0.42	0.37	0.70
Ireland	0.85	39	0.73	0.72	0.51	0.52	0.46	0.63	0.59	0.56	0.63	0.54	0.73	0.77
Israel	0.95	64	0.87	0.85	0.79	0.87	0.73	0.86	0.69	0.66	0.84	0.72	0.79	0.87
Italy	0.85	39	0.53	0.64	0.64	0.52	0.61	0.65	0.61	0.65	0.51	0.57	0.74	0.74
Kazakhstan	0.95	65	0.85	0.76	0.81	0.89	0.84	0.85	0.68	0.65	0.84	0.77	0.86	0.87
Kuwait	0.91	51	0.81	0.81	0.59	0.82	0.80	0.76	0.65	0.67	0.58	0.50	0.70	0.77
Latvia	0.94	62	0.90	0.86	0.83	0.80	0.85	0.87	0.73	0.66	0.81	0.28	0.80	0.88
Lithuania	0.88	43	0.72	0.67	0.51	0.65	0.54	0.68	0.73	0.66	0.65	0.57	0.65	0.79
Macao SAR	0.90	51	0.80	0.69	0.69	0.79	0.78	0.85	0.76	0.81	0.14	0.00	0.87	0.74
Malta	0.92	54	0.88	0.87	0.63	0.51	0.69	0.68	0.87	0.75	0.55	0.62	0.80	0.86
Morocco	0.84	40	0.35	0.51	0.42	0.71	0.29	0.77	0.75	0.76	0.68	0.78	0.77	0.49
Netherlands	0.77	29	0.42	0.24	0.53	0.46	0.42	0.50	0.54	0.48	0.53	0.69	0.69	0.78
New Zealand	0.88	46	0.73	0.61	0.59	0.63	0.51	0.72	0.65	0.73	0.64	0.72	0.74	0.79
Northern Ireland	0.90	48	0.75	0.78	0.71	0.73	0.69	0.59	0.78	0.73	0.51	0.59	0.65	0.77
Norway (5)	0.86	40	0.67	0.62	0.62	0.62	0.67	0.62	0.63	0.58	0.57	0.71	0.59	0.64
Oman	0.94	60	0.83	0.78	0.76	0.82	0.85	0.84	0.73	0.71	0.79	0.59	0.70	0.82
Poland	0.85	40	0.68	0.71	0.60	0.57	0.66	0.56	0.65	0.73	0.61	0.43	0.65	0.67
Portugal	0.90	50	0.75	0.74	0.74	0.71	0.73	0.67	0.63	0.62	0.59	0.73	0.75	0.77
Qatar	0.98	81	0.92	0.91	0.90	0.92	0.91	0.94	0.94	0.91	0.90	0.76	0.83	0.93
Russian Federation	0.91	53	0.77	0.76	0.58	0.81	0.67	0.73	0.70	0.69	0.68	0.67	0.81	0.79
Saudi Arabia	0.86	41	0.60	-	0.69	0.68	0.72	0.77	0.58	0.59	0.72	0.58	0.43	0.66
Singapore	0.96	71	0.92	0.85	0.85	0.89	0.86	0.78	0.89	0.89	0.59	0.71	0.87	0.91
Slovak Republic	0.85	38	0.64	0.74	0.58	0.61	0.58	0.59	0.71	0.64	0.56	0.57	0.49	0.69
Slovenia	0.83	37	0.66	0.62	0.40	0.44	0.47	0.38	0.76	0.70	0.66	0.59	0.72	0.72
South Africa	0.81	38	-0.15	-0.30	0.29	0.58	0.16	0.64	0.86	0.88	0.75	0.86	0.77	0.39
Spain	0.90	48	0.73	0.73	0.69	0.72	0.79	0.65	0.72	0.65	0.58	0.48	0.69	0.80
Sweden	0.86	41	0.72	0.62	0.56	0.51	0.56	0.66	0.68	0.62	0.60	0.74	0.63	0.74
Trinidad and Tobago	0.86	39	0.54	0.43	0.67	0.65	0.64	0.52	0.59	0.60	0.67	0.65	0.73	0.75
United Arab Emirates	0.96	71	0.87	0.87	0.85	0.90	0.87	0.89	0.86	0.81	0.82	0.62	0.86	0.87
United States	0.93	56	0.80	0.82	0.80	0.76	0.74	0.74	0.73	0.67	0.58	0.70	0.81	0.82
Benchmarking Participants														
Buenos Aires, Argentina	0.90	48	0.71	0.63	0.71	0.68	0.78	0.71	0.65	0.66	0.60	0.70	0.68	0.78
Ontario, Canada	0.92	54	0.69	0.74	0.65	0.59	0.64	0.76	0.81	0.75	0.68	0.76	0.85	0.84
Quebec, Canada	0.91	51	0.81	0.81	0.68	0.63	0.76	0.64	0.63	0.70	0.64	0.70	0.78	0.77
Denmark (3)	0.84	37	0.64	0.52	0.67	0.55	0.49	0.52	0.64	0.55	0.56	0.64	0.66	0.76
Norway (4)	0.85	40	0.67	0.61	0.64	0.60	0.68	0.63	0.61	0.58	0.59	0.70	0.60	0.63
Moscow City, Russian Fed.	0.94	62	0.85	0.78	0.70	0.87	0.83	0.80	0.71	0.73	0.77	0.60	0.85	0.90
Eng/Afr/Zulu - RSA (5)	0.83	36	0.13	0.08	0.41	0.65	0.35	0.65	0.76	0.73	0.80	0.79	0.70	0.59
Andalusia, Spain	0.83	35	0.56	0.56	0.61	0.66	0.70	0.51	0.70	0.56	0.47	0.52	0.58	0.67
Madrid, Spain	0.88	45	0.70	0.71	0.63	0.62	0.60	0.70	0.69	0.71	0.60	0.65	0.69	0.71
Abu Dhabi, UAE	0.96	67	0.85	0.84	0.82	0.87	0.86	0.88	0.84	0.79	0.79	0.53	0.86	0.83
Dubai, UAE	0.98	78	0.90	0.90	0.88	0.91	0.86	0.89	0.92	0.87	0.85	0.78	0.90	0.93

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Instruction Affected by Reading Resource Shortages* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.13	0.02	0.02
Austria	0.04	0.00	0.01
Azerbaijan	-0.13	0.02	0.02
Bahrain	0.13	0.02	0.02
Belgium (Flemish)	-0.01	0.00	0.00
Belgium (French)	0.06	0.00	0.00
Bulgaria	0.15	0.02	0.02
Canada	0.00	0.00	0.00
Chile	0.12	0.01	0.02
Chinese Taipei	0.07	0.01	0.00
Czech Republic	-0.01	0.00	0.00
Denmark	0.06	0.00	0.00
Egypt	0.12	0.01	0.01
England	0.05	0.00	0.00
Finland	0.04	0.00	0.00
France	0.08	0.01	0.00
Georgia	0.11	0.01	0.01
Germany	0.12	0.01	0.02
Hong Kong SAR	0.11	0.01	0.01
Hungary	-0.01	0.00	0.00
Iran, Islamic Rep. of	0.11	0.01	0.02
Ireland	0.05	0.00	0.00
Israel	0.29	0.08	0.07
Italy	0.01	0.00	0.00
Kazakhstan	0.04	0.00	0.01
Kuwait	0.03	0.00	0.01
Latvia	-0.02	0.00	0.00
Lithuania	0.05	0.00	0.00
Macao SAR	-0.13	0.02	0.01
Malta	0.06	0.00	0.00
Morocco	-0.07	0.01	0.01
Netherlands	0.04	0.00	0.00
New Zealand	0.13	0.02	0.01
Northern Ireland	-0.01	0.00	0.00
Norway (5)	0.09	0.01	0.01
Oman	0.07	0.00	0.01
Poland	0.05	0.00	0.00
Portugal	0.15	0.02	0.01
Qatar	0.24	0.06	0.05
Russian Federation	0.09	0.01	0.00
Saudi Arabia	-0.13	0.02	0.01
Singapore	-0.06	0.00	0.01
Slovak Republic	0.01	0.00	0.00
Slovenia	0.03	0.00	0.00
South Africa	0.19	0.04	0.03
Spain	0.18	0.03	0.05
Sweden	0.13	0.02	0.02
Trinidad and Tobago	0.11	0.01	0.00
United Arab Emirates	0.28	0.08	0.09
United States	-0.04	0.00	0.02
International Median	0.06	0.01	0.01
Benchmarking Participants			
Buenos Aires, Argentina	0.19	0.03	0.03
Ontario, Canada	-0.02	0.00	0.00
Quebec, Canada	0.03	0.00	0.00
Denmark (3)	0.01	0.00	0.00
Norway (4)	0.09	0.01	0.00
Moscow City, Russian Fed.	-0.01	0.00	0.00
Eng/Afr/Zulu - RSA (5)	0.21	0.05	0.07
Andalusia, Spain	0.13	0.02	0.01
Madrid, Spain	0.08	0.01	0.01
Abu Dhabi, UAE	0.28	0.08	0.08
Dubai, UAE	0.23	0.05	0.05

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Parents Like Reading Scale

The *Parents Like Reading* (PLR) scale was created based on parents' responses to the nine items listed below.

Items in the PIRLS 2016 *Parents Like Reading* Scale

Please indicate how much you agree with the following statements about reading.

			Agree a lot	Agree a little	Disagree a little	Disagree a lot
T	ASBH12A*	1) I read only if I have to*	○	○	○	○
T	ASBH12B	2) I like talking about what I read with other people	○	○	○	○
T	ASBH12C	3) I like to spend my spare time reading	○	○	○	○
T	ASBH12D*	4) I read only if I need information*	○	○	○	○
T	ASBH12E	5) Reading is an important activity in my home	○	○	○	○
T	ASBH12F	6) I would like to have more time for reading	○	○	○	○
T	ASBH12G	7) I enjoy reading	○	○	○	○
	ASBH12H	8) Reading is one of my favorite hobbies	○	○	○	○

Very Much Like | 10.5 | Somewhat Like | 8.1 | Do Not Like

			Every day or almost every day	Once or twice a week	Once or twice a month	Never or almost never
T	ASBH11	When you are at home, how often do you read for your enjoyment?	○	○	○	○

Very Much Like | 10.5 | Somewhat Like | 8.1 | Do Not Like

*reverse coded

T Trend item—item was included in the same scale in PIRLS 2011 and was used for linking the PIRLS 2011 and PIRLS 2016 scales.

Item Parameters for the PIRLS 2016 *Parents Like Reading* Scale

Items	delta	tau_1	tau_2	tau_3	Infit
* ASBH12A	0.44643	-0.55854	0.31019	0.24835	1.32
ASBH12B	-0.10611	-1.00923	-0.53256	1.54179	1.24
ASBH12C	0.06094	-0.99352	-0.30069	1.29421	0.81
* ASBH12D	0.69129	-0.76655	0.16445	0.60210	1.24
ASBH12E	-0.11197	-1.34252	-0.17347	1.51599	0.98
ASBH12F	-0.56903	-0.81012	-0.19368	1.00380	1.06
ASBH12G	-0.77034	-0.57881	-0.39616	0.97497	0.72
ASBH12H	0.15110	-0.96505	-0.17289	1.13794	0.81
ASBH11	0.20769	-0.67165	-0.49178	1.16343	1.13

* Reverse coded

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Parents Like Reading* Scale

Scale Transformation Constants	
A = 8.166833	Transformed Scale Score = 8.166833 + 1.409138 • Logit Scale Score
B = 1.409138	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Parents Like Reading* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	2.72380	
1	4.25018	
2	4.97008	
3	5.45813	
4	5.84059	
5	6.16163	
6	6.44443	
7	6.70075	
8	6.94217	
9	7.17109	
10	7.39134	
11	7.60584	
12	7.81680	
13	8.02605	8.1
14	8.23505	
15	8.44528	
16	8.65831	
17	8.87608	
18	9.10113	
19	9.33684	
20	9.58564	
21	9.85661	
22	10.15845	
23	10.50626	10.5
24	10.92449	
25	11.46501	
26	12.25570	
27	13.87824	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Parents Like Reading* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBH12A*	ASBH12B	ASBH12C	ASBH12D*	ASBH12E	ASBH12F	ASBH12G	ASBH12H	ASBH11
Australia	0.89	56	0.76	0.59	0.84	0.75	0.72	0.62	0.84	0.86	0.67
Austria	0.90	57	0.72	0.57	0.87	0.76	0.79	0.63	0.85	0.84	0.70
Azerbaijan	0.83	52	-	0.55	0.77	-	0.76	0.65	0.84	0.84	0.57
Bahrain	0.81	43	0.43	0.47	0.79	0.50	0.70	0.68	0.81	0.83	0.57
Belgium (Flemish)	0.90	57	0.75	0.50	0.87	0.75	0.75	0.67	0.88	0.86	0.70
Belgium (French)	0.90	55	0.70	0.52	0.85	0.72	0.72	0.71	0.86	0.86	0.70
Bulgaria	0.91	60	0.62	0.71	0.89	0.48	0.82	0.84	0.90	0.88	0.73
Canada	0.89	54	0.73	0.56	0.84	0.73	0.70	0.64	0.84	0.85	0.69
Chile	0.87	50	0.63	0.60	0.83	0.59	0.69	0.70	0.80	0.83	0.63
Chinese Taipei	0.86	49	0.57	0.48	0.77	0.54	0.76	0.77	0.86	0.86	0.56
Czech Republic	0.90	57	0.73	0.63	0.86	0.75	0.71	0.67	0.86	0.89	0.66
Denmark	0.90	57	0.74	0.60	0.88	0.78	0.76	0.60	0.85	0.83	0.70
Egypt	0.79	55	-0.25	0.78	0.83	-0.41	0.85	0.86	0.91	0.87	0.58
England	-	-	-	-	-	-	-	-	-	-	-
Finland	0.91	58	0.73	0.57	0.87	0.76	0.82	0.65	0.86	0.87	0.67
France	0.88	51	0.69	0.53	0.82	0.70	0.65	0.60	0.82	0.84	0.71
Georgia	0.78	43	0.24	0.56	0.81	0.26	0.68	0.75	0.85	0.81	0.61
Germany	0.89	54	0.75	0.57	0.87	0.66	0.70	0.61	0.86	0.85	0.68
Hong Kong SAR	0.85	48	0.46	0.52	0.78	0.41	0.75	0.78	0.87	0.87	0.61
Hungary	0.90	56	0.70	0.60	0.84	0.71	0.74	0.74	0.83	0.86	0.65
Iran, Islamic Rep. of	0.81	44	0.39	0.55	0.79	0.28	0.71	0.78	0.82	0.82	0.60
Ireland	0.88	53	0.74	0.55	0.83	0.74	0.69	0.57	0.84	0.86	0.69
Israel	0.85	47	0.65	0.50	0.81	0.65	0.68	0.57	0.79	0.84	0.62
Italy	0.88	52	0.69	0.55	0.82	0.69	0.70	0.67	0.81	0.83	0.67
Kazakhstan	0.72	39	0.11	0.51	0.74	0.20	0.70	0.73	0.77	0.82	0.60
Kuwait	0.84	47	0.52	0.49	0.80	0.58	0.75	0.70	0.81	0.84	0.54
Latvia	0.87	51	0.68	0.37	0.84	0.72	0.69	0.68	0.84	0.85	0.61
Lithuania	0.89	54	0.68	0.56	0.85	0.65	0.75	0.70	0.85	0.86	0.64
Macao SAR	0.83	46	0.38	0.56	0.76	0.35	0.76	0.78	0.87	0.87	0.55
Malta	0.86	49	0.68	0.49	0.82	0.72	0.57	0.64	0.82	0.86	0.62
Morocco	0.84	55	-0.04	0.81	0.86	-0.19	0.86	0.82	0.90	0.88	0.69
Netherlands	0.89	56	0.81	0.54	0.87	0.76	0.61	0.55	0.87	0.87	0.74
New Zealand	0.89	54	0.73	0.53	0.85	0.72	0.73	0.61	0.84	0.87	0.68
Northern Ireland	0.91	59	0.82	0.66	0.86	0.80	0.71	0.58	0.82	0.88	0.75
Norway (5)	0.89	53	0.73	0.52	0.85	0.74	0.74	0.60	0.83	0.83	0.66
Oman	0.78	41	0.24	0.55	0.77	0.34	0.69	0.69	0.80	0.81	0.58
Poland	0.88	53	0.70	0.61	0.83	0.71	0.75	0.62	0.81	0.80	0.67
Portugal	0.87	50	0.68	0.55	0.80	0.67	0.72	0.64	0.79	0.84	0.61
Qatar	0.81	43	0.44	0.44	0.79	0.46	0.70	0.68	0.82	0.83	0.54
Russian Federation	0.86	49	0.61	0.51	0.82	0.64	0.65	0.68	0.83	0.85	0.61
Saudi Arabia	0.82	44	0.38	0.55	0.79	0.34	0.74	0.69	0.83	0.83	0.61
Singapore	0.85	49	0.58	0.43	0.82	0.54	0.72	0.75	0.86	0.87	0.59
Slovak Republic	0.90	57	0.74	0.62	0.87	0.71	0.72	0.68	0.87	0.89	0.68
Slovenia	0.88	53	0.74	0.54	0.85	0.63	0.70	0.69	0.87	0.83	0.59
South Africa	0.75	41	0.03	0.66	0.76	0.05	0.70	0.71	0.82	0.80	0.61
Spain	0.89	54	0.71	0.59	0.83	0.73	0.67	0.67	0.83	0.85	0.70
Sweden	0.89	55	0.78	0.58	0.88	0.77	0.68	0.56	0.85	0.82	0.67
Trinidad and Tobago	0.82	44	0.63	0.53	0.80	0.63	0.60	0.45	0.79	0.83	0.57
United Arab Emirates	0.80	42	0.41	0.49	0.78	0.41	0.72	0.68	0.81	0.82	0.53
United States	-	-	-	-	-	-	-	-	-	-	-
Benchmarking Participants											
Buenos Aires, Argentina	0.84	46	0.62	0.51	0.82	0.62	0.62	0.66	0.78	0.83	0.57
Ontario, Canada	0.88	53	0.73	0.53	0.83	0.73	0.69	0.64	0.83	0.85	0.67
Quebec, Canada	0.90	55	0.74	0.63	0.85	0.72	0.70	0.62	0.84	0.85	0.70
Denmark (3)	0.89	55	0.74	0.58	0.87	0.77	0.75	0.61	0.84	0.82	0.66
Norway (4)	0.88	53	0.72	0.53	0.85	0.73	0.75	0.60	0.83	0.83	0.63
Moscow City, Russian Fed.	0.86	49	0.65	0.47	0.82	0.69	0.67	0.63	0.83	0.83	0.61
Eng/Afr/Zulu - RSA (5)	0.78	41	0.28	0.62	0.77	0.29	0.69	0.60	0.81	0.80	0.63
Andalusia, Spain	0.89	55	0.71	0.59	0.83	0.73	0.69	0.67	0.85	0.86	0.70
Madrid, Spain	0.87	51	0.69	0.55	0.82	0.74	0.68	0.58	0.81	0.85	0.65
Abu Dhabi, UAE	0.79	41	0.39	0.50	0.78	0.37	0.73	0.66	0.81	0.82	0.49
Dubai, UAE	0.82	44	0.54	0.44	0.78	0.56	0.71	0.66	0.81	0.82	0.57

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

A dash (–) indicates comparable data not available.
*Reverse coded

Relationship Between the PIRLS 2016 *Parents Like Reading* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.24	0.06	0.05
Austria	0.30	0.09	0.08
Azerbaijan	0.11	0.01	0.01
Bahrain	0.19	0.04	0.03
Belgium (Flemish)	0.23	0.05	0.05
Belgium (French)	0.28	0.08	0.08
Bulgaria	0.41	0.17	0.16
Canada	0.21	0.04	0.04
Chile	0.23	0.05	0.05
Chinese Taipei	0.19	0.04	0.04
Czech Republic	0.28	0.08	0.07
Denmark	0.21	0.04	0.04
Egypt	0.30	0.09	0.08
England	-	-	-
Finland	0.25	0.06	0.06
France	0.25	0.06	0.05
Georgia	0.20	0.04	0.04
Germany	0.35	0.12	0.11
Hong Kong SAR	0.09	0.01	0.01
Hungary	0.35	0.12	0.11
Iran, Islamic Rep. of	0.25	0.06	0.05
Ireland	0.26	0.07	0.05
Israel	0.18	0.03	0.03
Italy	0.22	0.05	0.04
Kazakhstan	0.09	0.01	0.01
Kuwait	0.17	0.03	0.02
Latvia	0.22	0.05	0.05
Lithuania	0.24	0.06	0.05
Macao SAR	0.13	0.02	0.01
Malta	0.14	0.02	0.02
Morocco	0.24	0.06	0.05
Netherlands	0.26	0.07	0.06
New Zealand	0.29	0.09	0.07
Northern Ireland	0.18	0.03	0.03
Norway (5)	0.25	0.06	0.05
Oman	0.20	0.04	0.03
Poland	0.21	0.05	0.04
Portugal	0.21	0.05	0.04
Qatar	0.21	0.05	0.04
Russian Federation	0.22	0.05	0.04
Saudi Arabia	0.14	0.02	0.02
Singapore	0.20	0.04	0.04
Slovak Republic	0.38	0.15	0.11
Slovenia	0.26	0.07	0.06
South Africa	0.17	0.03	0.03
Spain	0.20	0.04	0.04
Sweden	0.26	0.07	0.06
Trinidad and Tobago	0.12	0.02	0.02
United Arab Emirates	0.22	0.05	0.04
United States	-	-	-
International Median	0.22	0.05	0.04
Benchmarking Participants			
Buenos Aires, Argentina	0.24	0.06	0.06
Ontario, Canada	0.20	0.04	0.04
Quebec, Canada	0.17	0.03	0.03
Denmark (3)	0.18	0.03	0.03
Norway (4)	0.24	0.06	0.05
Moscow City, Russian Fed.	0.21	0.04	0.04
Eng/Afr/Zulu - RSA (5)	0.16	0.02	0.03
Andalusia, Spain	0.22	0.05	0.05
Madrid, Spain	0.18	0.03	0.03
Abu Dhabi, UAE	0.22	0.05	0.04
Dubai, UAE	0.23	0.05	0.05

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Parents' Perceptions of Their Child's School Scale

The *Parents' Perceptions of Their Child's School* (PCS) scale was created based on parents' responses to the six statements described below.

Items in the PIRLS 2016 *Parents' Perceptions of Their Child's School* Scale[1]

What do you think of your child's school?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBH09A	1) My child's school does a good job including me in my child's education	○	○	○	○
ASBH09B	2) My child's school provides a safe environment	○	○	○	○
ASBH09C	3) My child's school cares about my child's progress in school	○	○	○	○
ASBH09D	4) My child's school does a good job informing me of his/her progress	○	○	○	○
ASBH09E	5) My child's school promotes high academic standards	○	○	○	○
ASBH09F	6) My child's school does a good job in helping him/her become better in reading	○	○	○	○

Very Satisfied | 9.5 | **Somewhat Satisfied** | 6.3 | **Less than Satisfied**

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the PIRLS 2016 *Parents' Perceptions of Their Child's School* Scale

Item	delta	tau_1	tau_2	Infit
ASBH09A	0.08775	-1.64719	1.64719	1.02
ASBH09B	-0.56014	-1.57499	1.57499	1.22
ASBH09C	-0.46209	-1.61352	1.61352	0.82
ASBH09D	0.03829	-1.40332	1.40332	0.96
ASBH09E	0.90366	-1.57716	1.57716	1.14
ASBH09F	-0.00747	-1.46353	1.46353	1.00

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Parents' Perceptions of Their Child's School* Scale

Scale Transformation Constants	
A = 7.908785	Transformed Scale Score = 7.908785 + 1.026352 • Logit Scale Score
B = 1.026352	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Parents' Perceptions of Their Child's School* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.57843	
1	4.89714	
2	5.63616	
3	6.23689	6.3
4	6.79025	
5	7.34568	
6	7.91434	
7	8.47752	
8	9.02158	
9	9.57101	9.5
10	10.16827	
11	10.92169	
12	12.26356	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Parents' Perceptions of Their Child's School* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			ASBH09A	ASBH09B	ASBH09C	ASBH09D	ASBH09E	ASBH09F
Australia	0.88	63	0.83	0.60	0.87	0.84	0.80	0.80
Austria	0.86	59	0.79	0.62	0.83	0.83	0.75	0.78
Azerbaijan	0.83	55	0.68	0.76	0.80	0.72	0.72	0.75
Bahrain	0.87	60	0.77	0.70	0.82	0.79	0.80	0.76
Belgium (Flemish)	0.85	58	0.81	0.68	0.84	0.85	0.63	0.73
Belgium (French)	0.87	60	0.79	0.72	0.84	0.82	0.70	0.77
Bulgaria	0.87	62	0.75	0.75	0.83	0.80	0.80	0.79
Canada	0.88	62	0.82	0.61	0.85	0.84	0.80	0.78
Chile	0.89	64	0.80	0.72	0.86	0.84	0.79	0.79
Chinese Taipei	0.88	63	0.78	0.77	0.88	0.87	0.67	0.78
Czech Republic	0.85	58	0.72	0.62	0.83	0.81	0.77	0.78
Denmark	0.89	64	0.80	0.71	0.85	0.85	0.82	0.78
Egypt	0.87	62	0.79	0.75	0.83	0.80	0.75	0.79
England	-	-	-	-	-	-	-	-
Finland	0.82	53	0.74	0.64	0.83	0.76	0.63	0.73
France	0.88	63	0.79	0.72	0.85	0.82	0.80	0.78
Georgia	0.84	56	0.71	0.64	0.82	0.76	0.79	0.77
Germany	0.87	60	0.82	0.64	0.83	0.82	0.76	0.76
Hong Kong SAR	0.85	57	0.80	0.62	0.83	0.83	0.66	0.77
Hungary	0.86	58	0.78	0.66	0.81	0.80	0.75	0.78
Iran, Islamic Rep. of	0.83	55	0.73	0.68	0.80	0.79	0.66	0.77
Ireland	0.85	58	0.79	0.59	0.82	0.82	0.76	0.76
Israel	0.89	64	0.81	0.68	0.85	0.84	0.80	0.81
Italy	0.84	57	0.75	0.59	0.81	0.80	0.78	0.76
Kazakhstan	0.80	51	0.72	0.61	0.79	0.67	0.72	0.75
Kuwait	0.89	65	0.78	0.75	0.86	0.82	0.82	0.81
Latvia	0.84	55	0.64	0.66	0.83	0.71	0.83	0.78
Lithuania	0.81	52	0.69	0.69	0.81	0.73	0.67	0.73
Macao SAR	0.85	57	0.77	0.65	0.81	0.81	0.73	0.75
Malta	0.75	46	0.59	0.62	0.77	0.70	0.72	0.66
Morocco	0.82	53	0.74	0.68	0.80	0.72	0.65	0.79
Netherlands	0.83	56	0.80	0.67	0.85	0.84	0.53	0.74
New Zealand	0.88	62	0.82	0.62	0.85	0.83	0.79	0.81
Northern Ireland	0.86	61	0.83	0.63	0.85	0.81	0.77	0.77
Norway (5)	0.86	59	0.78	0.63	0.84	0.82	0.77	0.77
Oman	0.85	57	0.77	0.66	0.81	0.76	0.76	0.76
Poland	0.87	61	0.74	0.74	0.85	0.75	0.82	0.79
Portugal	0.85	59	0.79	0.66	0.84	0.77	0.77	0.76
Qatar	0.89	64	0.79	0.71	0.86	0.81	0.81	0.81
Russian Federation	0.86	59	0.63	0.71	0.83	0.75	0.85	0.83
Saudi Arabia	0.86	61	0.78	0.71	0.83	0.79	0.73	0.81
Singapore	0.86	60	0.79	0.66	0.84	0.81	0.76	0.76
Slovak Republic	0.87	61	0.73	0.66	0.85	0.81	0.83	0.80
Slovenia	0.84	56	0.69	0.68	0.82	0.82	0.72	0.76
South Africa	0.83	54	0.68	0.67	0.75	0.76	0.74	0.78
Spain	0.85	58	0.77	0.68	0.83	0.81	0.72	0.76
Sweden	0.90	67	0.81	0.74	0.87	0.83	0.84	0.80
Trinidad and Tobago	0.87	61	0.81	0.63	0.85	0.79	0.79	0.78
United Arab Emirates	0.89	64	0.80	0.70	0.85	0.83	0.83	0.79
United States	-	-	-	-	-	-	-	-
Benchmarking Participants								
Buenos Aires, Argentina	0.82	53	0.74	0.58	0.82	0.75	0.74	0.74
Ontario, Canada	0.88	62	0.82	0.59	0.85	0.83	0.82	0.78
Quebec, Canada	0.87	60	0.80	0.65	0.84	0.83	0.76	0.76
Denmark (3)	0.88	63	0.82	0.68	0.84	0.86	0.81	0.76
Norway (4)	0.86	60	0.79	0.65	0.83	0.81	0.79	0.75
Moscow City, Russian Fed.	0.84	56	0.54	0.69	0.82	0.73	0.85	0.82
Eng/Afr/Zulu - RSA (5)	0.84	56	0.74	0.64	0.79	0.78	0.77	0.77
Andalusia, Spain	0.85	59	0.78	0.67	0.83	0.80	0.72	0.78
Madrid, Spain	0.86	59	0.80	0.65	0.84	0.81	0.70	0.77
Abu Dhabi, UAE	0.90	67	0.80	0.72	0.87	0.85	0.85	0.80
Dubai, UAE	0.87	61	0.79	0.64	0.85	0.81	0.80	0.77

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Parents' Perceptions of Their Child's School* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.02	0.00	0.00
Austria	-0.03	0.00	0.00
Azerbaijan	0.12	0.01	0.01
Bahrain	0.19	0.04	0.04
Belgium (Flemish)	-0.06	0.00	0.00
Belgium (French)	0.00	0.00	0.00
Bulgaria	-0.06	0.00	0.00
Canada	0.01	0.00	0.00
Chile	0.03	0.00	0.00
Chinese Taipei	-0.06	0.00	0.00
Czech Republic	-0.11	0.01	0.01
Denmark	0.07	0.01	0.01
Egypt	0.08	0.01	0.01
England	-	-	-
Finland	-0.02	0.00	0.00
France	-0.01	0.00	0.00
Georgia	0.01	0.00	0.00
Germany	0.06	0.00	0.01
Hong Kong SAR	0.11	0.01	0.01
Hungary	-0.01	0.00	0.00
Iran, Islamic Rep. of	-0.01	0.00	0.00
Ireland	0.01	0.00	0.00
Israel	-0.13	0.02	0.01
Italy	0.03	0.00	0.00
Kazakhstan	-0.02	0.00	0.00
Kuwait	0.12	0.02	0.01
Latvia	-0.04	0.00	0.00
Lithuania	0.02	0.00	0.00
Macao SAR	0.12	0.01	0.01
Malta	0.08	0.01	0.00
Morocco	0.23	0.05	0.04
Netherlands	0.06	0.00	0.01
New Zealand	0.02	0.00	0.00
Northern Ireland	0.07	0.00	0.01
Norway (5)	0.07	0.01	0.01
Oman	0.12	0.01	0.02
Poland	-0.06	0.00	0.00
Portugal	0.02	0.00	0.00
Qatar	0.13	0.02	0.02
Russian Federation	-0.02	0.00	0.00
Saudi Arabia	0.12	0.01	0.02
Singapore	0.06	0.00	0.00
Slovak Republic	-0.08	0.01	0.00
Slovenia	-0.08	0.01	0.01
South Africa	0.13	0.02	0.02
Spain	0.02	0.00	0.00
Sweden	0.05	0.00	0.01
Trinidad and Tobago	0.15	0.02	0.01
United Arab Emirates	0.17	0.03	0.03
United States	-	-	-
International Median	0.02	0.00	0.00
Benchmarking Participants			
Buenos Aires, Argentina	-0.02	0.00	0.00
Ontario, Canada	0.01	0.00	0.00
Quebec, Canada	-0.05	0.00	0.01
Denmark (3)	0.07	0.00	0.01
Norway (4)	-0.01	0.00	0.00
Moscow City, Russian Fed.	-0.03	0.00	0.00
Eng/Afr/Zulu - RSA (5)	0.12	0.02	0.01
Andalusia, Spain	-0.04	0.00	0.00
Madrid, Spain	0.04	0.00	0.00
Abu Dhabi, UAE	0.15	0.02	0.03
Dubai, UAE	0.18	0.03	0.03

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Safe and Orderly School Scale

The *Safe and Orderly School* (SOS) scale was created based on teachers' degree of agreement with the eight statements described below.

Items in the PIRLS 2016 *Safe and Orderly School* Scale[1]

Thinking about your current school, indicate the extent to which you agree or disagree with each of the following statements.

	Variable	Statement	Agree a lot	Agree a little	Disagree a little	Disagree a lot
T	ATBG08A	1) This school is located in a safe neighborhood	○	○	○	○
T	ATBG08B	2) I feel safe at this school	○	○	○	○
T	ATBG08C	3) This school's security policies and practices are sufficient	○	○	○	○
T	ATBG08D	4) The students behave in an orderly manner	○	○	○	○
T	ATBG08E	5) The students are respectful of the teachers	○	○	○	○
	ATBG08F	6) The students respect school property	○	○	○	○
	ATBG08G	7) This school has clear rules about student conduct	○	○	○	○
	ATBG08H	8) This school's rules are enforced in a fair and consistent manner	○	○	○	○

Very Safe and Orderly | 9.9 | Somewhat Safe and Orderly | 6.6 | Less than Safe and Orderly

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Disagree a little" and "Disagree a lot" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

T Trend item—item was included in the same scale in PIRLS 2011 and was used for linking the PIRLS 2011 and PIRLS 2016 scales.

Item Parameters for the PIRLS 2016 *Safe and Orderly School* Scale

Item	delta	tau_1	tau_2	Infit
ATBG08A	-0.33477	-0.95246	0.95246	1.44
ATBG08B	-1.33999	-1.18806	1.18806	1.00
ATBG08C	-0.43632	-1.32751	1.32751	1.11
ATBG08D	0.99788	-1.94721	1.94721	0.86
ATBG08E	0.46648	-1.96220	1.96220	0.88
ATBG08F	1.16567	-1.93026	1.93026	0.90
ATBG08G	-0.61716	-1.39812	1.39812	1.05
ATBG08H	0.09821	-1.55388	1.55388	1.06

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Safe and Orderly School* Scale

Scale Transformation Constants	
A = 8.265816	Transformed Scale Score = 8.265816 + 1.015430 • Logit Scale Score
B = 1.015430	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Safe and Orderly School* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.68014	
1	4.94412	
2	5.61774	
3	6.12515	
4	6.55836	6.6
5	6.95570	
6	7.33778	
7	7.71926	
8	8.10772	
9	8.51407	
10	8.94594	
11	9.41153	
12	9.92288	9.9
13	10.49280	
14	11.14276	
15	11.94193	
16	13.31150	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Safe and Orderly School* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			ATBG08A	ATBG08B	ATBG08C	ATBG08D	ATBG08E	ATBG08F	ATBG08G	ATBG08H
Australia	0.88	55	0.65	0.69	0.63	0.84	0.84	0.83	0.64	0.77
Austria	0.82	45	0.56	0.67	0.61	0.73	0.77	0.71	0.62	0.70
Azerbaijan	0.72	36	0.40	0.39	0.60	0.68	0.58	0.67	0.67	0.72
Bahrain	0.89	56	0.60	0.61	0.68	0.84	0.84	0.82	0.81	0.75
Belgium (Flemish)	0.83	46	0.61	0.52	0.62	0.74	0.74	0.73	0.73	0.68
Belgium (French)	0.85	50	0.65	0.65	0.66	0.81	0.76	0.76	0.62	0.71
Bulgaria	0.84	48	0.48	0.70	0.74	0.81	0.73	0.73	0.64	0.63
Canada	0.88	55	0.55	0.63	0.69	0.82	0.83	0.80	0.77	0.80
Chile	0.88	56	0.69	0.73	0.72	0.78	0.83	0.75	0.69	0.77
Chinese Taipei	0.86	51	0.53	0.71	0.77	0.78	0.79	0.73	0.71	0.64
Czech Republic	0.83	46	0.52	0.64	0.65	0.74	0.76	0.70	0.72	0.68
Denmark	0.86	51	0.65	0.47	0.68	0.83	0.77	0.80	0.71	0.74
Egypt	0.85	48	0.50	0.58	0.72	0.71	0.69	0.78	0.76	0.75
England	0.77	42	0.50	0.51	0.44	0.84	0.81	0.77	0.53	0.62
Finland	0.88	54	0.56	0.71	0.62	0.82	0.80	0.82	0.72	0.80
France	0.85	49	0.73	0.70	0.69	0.74	0.75	0.73	0.58	0.65
Georgia	0.71	33	0.48	0.55	0.64	0.65	0.64	0.68	0.47	0.46
Germany	0.84	49	0.71	0.57	0.49	0.80	0.78	0.79	0.66	0.71
Hong Kong SAR	0.84	48	0.54	0.67	0.63	0.70	0.77	0.76	0.76	0.70
Hungary	0.84	48	0.46	0.70	0.76	0.76	0.71	0.76	0.64	0.71
Iran, Islamic Rep. of	0.84	48	0.65	0.62	0.70	0.70	0.69	0.66	0.75	0.75
Ireland	0.86	52	0.46	0.61	0.70	0.85	0.84	0.72	0.74	0.76
Israel	0.88	55	0.62	0.71	0.70	0.83	0.75	0.73	0.78	0.81
Italy	0.84	48	0.62	0.70	0.66	0.71	0.70	0.75	0.70	0.68
Kazakhstan	0.74	39	0.59	0.44	0.61	0.66	0.71	0.69	0.69	0.55
Kuwait	0.81	44	0.20	0.43	0.46	0.79	0.85	0.79	0.69	0.81
Latvia	0.78	40	0.44	0.55	0.60	0.74	0.69	0.65	0.66	0.70
Lithuania	0.82	44	0.44	0.74	0.69	0.65	0.67	0.74	0.59	0.77
Macao SAR	0.84	48	0.56	0.58	0.58	0.78	0.77	0.75	0.76	0.68
Malta	0.88	54	0.47	0.69	0.78	0.82	0.80	0.81	0.69	0.77
Morocco	0.89	57	0.73	0.78	0.79	0.73	0.75	0.78	0.76	0.74
Netherlands	0.80	43	0.44	0.58	0.55	0.83	0.77	0.74	0.48	0.73
New Zealand	0.87	54	0.70	0.73	0.67	0.81	0.81	0.69	0.71	0.75
Northern Ireland	0.84	48	0.48	0.32	0.64	0.87	0.76	0.80	0.75	0.76
Norway (5)	0.83	46	0.45	0.48	0.58	0.77	0.79	0.77	0.76	0.70
Oman	0.82	45	0.51	0.53	0.68	0.73	0.71	0.68	0.70	0.76
Poland	0.80	43	0.39	0.54	0.48	0.79	0.81	0.74	0.67	0.70
Portugal	0.86	52	0.53	0.68	0.63	0.80	0.79	0.75	0.75	0.76
Qatar	0.83	48	0.43	0.52	0.60	0.75	0.79	0.82	0.78	0.74
Russian Federation	0.83	46	0.56	0.66	0.62	0.76	0.76	0.77	0.53	0.75
Saudi Arabia	0.84	48	0.52	0.62	0.68	0.77	0.70	0.78	0.73	0.72
Singapore	0.87	53	0.51	0.60	0.63	0.82	0.83	0.82	0.76	0.81
Slovak Republic	0.86	51	0.52	0.72	0.66	0.80	0.78	0.74	0.74	0.75
Slovenia	0.88	55	0.62	0.76	0.77	0.73	0.80	0.74	0.73	0.75
South Africa	0.87	54	0.66	0.78	0.66	0.83	0.81	0.78	0.61	0.69
Spain	0.86	52	0.57	0.65	0.71	0.79	0.82	0.80	0.64	0.75
Sweden	0.86	52	0.54	0.69	0.69	0.81	0.86	0.80	0.56	0.77
Trinidad and Tobago	0.89	57	0.68	0.77	0.76	0.86	0.85	0.84	0.57	0.69
United Arab Emirates	0.87	53	0.39	0.55	0.61	0.80	0.84	0.84	0.80	0.83
United States	0.90	58	0.71	0.72	0.76	0.84	0.85	0.88	0.64	0.67
Benchmarking Participants										
Buenos Aires, Argentina	0.87	54	0.51	0.76	0.72	0.79	0.78	0.80	0.72	0.74
Ontario, Canada	0.89	58	0.55	0.64	0.71	0.85	0.84	0.80	0.79	0.84
Quebec, Canada	0.82	45	0.53	0.49	0.61	0.78	0.77	0.73	0.70	0.70
Denmark (3)	0.83	45	0.41	0.33	0.54	0.83	0.81	0.79	0.72	0.76
Norway (4)	0.82	45	0.49	0.48	0.58	0.77	0.73	0.77	0.66	0.80
Moscow City, Russian Fed.	0.84	48	0.63	0.69	0.69	0.68	0.75	0.75	0.57	0.76
Eng/Afr/Zulu - RSA (5)	0.88	56	0.67	0.60	0.80	0.87	0.79	0.81	0.67	0.74
Andalusia, Spain	0.87	54	0.70	0.76	0.69	0.78	0.78	0.79	0.60	0.74
Madrid, Spain	0.85	49	0.61	0.65	0.74	0.82	0.78	0.82	0.51	0.62
Abu Dhabi, UAE	0.85	49	0.41	0.52	0.50	0.81	0.81	0.83	0.79	0.79
Dubai, UAE	0.87	54	0.41	0.51	0.65	0.83	0.85	0.85	0.78	0.83

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Safe and Orderly School* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.21	0.04	0.02
Austria	0.10	0.01	0.01
Azerbaijan	0.08	0.01	0.00
Bahrain	0.24	0.06	0.04
Belgium (Flemish)	0.14	0.02	0.02
Belgium (French)	0.17	0.03	0.03
Bulgaria	0.23	0.05	0.05
Canada	0.09	0.01	0.01
Chile	0.20	0.04	0.06
Chinese Taipei	-0.03	0.00	0.01
Czech Republic	0.09	0.01	0.02
Denmark	0.09	0.01	0.01
Egypt	0.10	0.01	0.01
England	0.11	0.01	0.02
Finland	0.03	0.00	0.00
France	0.13	0.02	0.02
Georgia	0.08	0.01	0.00
Germany	0.25	0.06	0.07
Hong Kong SAR	0.04	0.00	0.00
Hungary	0.16	0.02	0.03
Iran, Islamic Rep. of	0.08	0.01	0.00
Ireland	0.11	0.01	0.01
Israel	0.06	0.00	0.01
Italy	0.09	0.01	0.01
Kazakhstan	-0.02	0.00	0.00
Kuwait	0.13	0.02	0.00
Latvia	0.04	0.00	0.00
Lithuania	0.09	0.01	0.03
Macao SAR	0.10	0.01	0.01
Malta	0.08	0.01	0.01
Morocco	0.26	0.07	0.06
Netherlands	0.18	0.03	0.04
New Zealand	0.21	0.05	0.04
Northern Ireland	0.09	0.01	0.01
Norway (5)	0.08	0.01	0.01
Oman	0.10	0.01	0.00
Poland	-0.04	0.00	0.00
Portugal	0.13	0.02	0.01
Qatar	0.07	0.00	0.00
Russian Federation	0.04	0.00	0.00
Saudi Arabia	0.17	0.03	0.02
Singapore	0.08	0.01	0.00
Slovak Republic	0.13	0.02	0.03
Slovenia	-0.01	0.00	0.00
South Africa	0.00	0.00	0.00
Spain	0.16	0.02	0.01
Sweden	0.16	0.03	0.03
Trinidad and Tobago	0.12	0.02	0.01
United Arab Emirates	0.26	0.07	0.06
United States	0.23	0.05	0.05
International Median	0.10	0.01	0.01
Benchmarking Participants			
Buenos Aires, Argentina	0.20	0.04	0.04
Ontario, Canada	0.11	0.01	0.02
Quebec, Canada	0.01	0.00	0.00
Denmark (3)	0.06	0.00	0.00
Norway (4)	0.01	0.00	0.00
Moscow City, Russian Fed.	0.02	0.00	0.00
Eng/Afr/Zulu - RSA (5)	0.03	0.00	0.01
Andalusia, Spain	0.22	0.05	0.04
Madrid, Spain	0.14	0.02	0.01
Abu Dhabi, UAE	0.19	0.04	0.03
Dubai, UAE	0.26	0.07	0.06

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

School Discipline Scale

The *School Discipline* (DAS) scale was created based on principals' responses concerning the ten potential school problems described below.

Items in the PIRLS 2016 *School Discipline* Scale

To what degree is each of the following a problem among fourth grade students in your school?

			Not a problem	Minor problem	Moderate problem	Serious problem
T	ACBG14A	1) Arriving late at school	○	○	○	○
T	ACBG14B	2) Absenteeism (i.e., unjustified absences)	○	○	○	○
T	ACBG14C	3) Classroom disturbance	○	○	○	○
T	ACBG14D	4) Cheating	○	○	○	○
T	ACBG14E	5) Profanity	○	○	○	○
T	ACBG14F	6) Vandalism	○	○	○	○
T	ACBG14G	7) Theft	○	○	○	○
T	ACBG14H	8) Intimidation or verbal abuse among students (including texting, emailing, etc.)	○	○	○	○
T	ACBG14I	9) Physical fights among students	○	○	○	○
T	ACBG14J	10) Intimidation or verbal abuse of teachers or staff (including texting, emailing, etc.)	○	○	○	○

Hardly Any Problems | 9.9 | **Minor Problems** | 7.7 | **Moderate to Severe Problems**

T Trend item—item was included in the same scale in PIRLS 2011 and was used for linking the PIRLS 2011 and PIRLS 2016 scales.

Item Parameters for the PIRLS 2016 *School Discipline* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ACBG14A	-0.12195	-2.34683	-0.42737	2.77420	1.24
ACBG14B	0.12937	-1.53131	-0.45652	1.98783	1.16
ACBG14C	0.91425	-2.22837	-0.27536	2.50373	1.02
ACBG14D	-0.39571	-1.01520	-0.90302	1.91822	1.08
ACBG14E	0.57429	-1.73047	-0.41827	2.14874	0.91
ACBG14F	-0.35555	-0.41221	-0.73370	1.14591	0.75
ACBG14G	-0.55939	0.41486	-1.38128	0.96642	0.84
ACBG14H	0.25527	-1.55859	-0.58526	2.14385	0.93
ACBG14I	0.33890	-1.63156	-0.63374	2.26530	0.84
ACBG14J	-0.77948	-0.03064	-0.67247	0.70311	0.86

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *School Discipline* Scale

Scale Transformation Constants	
A = 7.915470	Transformed Scale Score = 7.915470 + 0.941833 • Logit Scale Score
B = 0.941833	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *School Discipline* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.78131	
1	4.85875	
2	5.36727	
3	5.70447	
4	5.96046	
5	6.16894	
6	6.34794	
7	6.50757	
8	6.65426	
9	6.79253	
10	6.92445	
11	7.05592	
12	7.18729	
13	7.32077	
14	7.45855	
15	7.60360	7.7
16	7.75631	
17	7.92096	
18	8.09978	
19	8.29514	
20	8.51076	
21	8.74775	
22	9.00813	
23	9.29258	
24	9.60146	
25	9.93429	9.9
26	10.29556	
27	10.69797	
28	11.17296	
29	11.80710	
30	12.98539	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *School Discipline* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Components Loadings for Each Item									
			ACBG14A	ACBG14B	ACBG14C	ACBG14D	ACBG14E	ACBG14F	ACBG14G	ACBG14H	ACBG14I	ACBG14J
Australia	0.88	51	0.58	0.59	0.72	0.63	0.84	0.75	0.76	0.77	0.76	0.73
Austria	0.86	45	0.48	0.56	0.76	0.62	0.83	0.77	0.63	0.70	0.74	0.53
Azerbaijan	0.95	70	0.50	0.73	0.86	0.80	0.84	0.93	0.92	0.88	0.87	0.94
Bahrain	0.96	74	0.74	0.82	0.78	0.90	0.87	0.91	0.91	0.88	0.84	0.92
Belgium (Flemish)	0.86	46	0.51	0.58	0.72	0.68	0.77	0.74	0.61	0.72	0.74	0.68
Belgium (French)	0.85	43	0.51	0.64	0.54	0.55	0.73	0.65	0.69	0.79	0.78	0.65
Bulgaria	0.93	64	0.64	0.80	0.81	0.82	0.73	0.81	0.85	0.82	0.84	0.86
Canada	0.88	50	0.62	0.61	0.71	0.64	0.77	0.74	0.74	0.74	0.77	0.69
Chile	0.92	59	0.57	0.60	0.73	0.65	0.83	0.87	0.86	0.86	0.82	0.79
Chinese Taipei	0.88	50	0.61	0.63	0.68	0.68	0.82	0.74	0.76	0.77	0.74	0.59
Czech Republic	0.83	41	0.48	0.49	0.64	0.69	0.75	0.66	0.65	0.73	0.56	0.67
Denmark	0.82	40	0.59	0.53	0.70	0.45	0.77	0.58	0.49	0.79	0.77	0.54
Egypt	0.93	61	0.61	0.63	0.75	0.80	0.86	0.85	0.86	0.85	0.75	0.79
England	0.77	34	0.67	0.67	0.63	0.14	0.71	0.48	0.36	0.67	0.74	0.51
Finland	0.79	35	0.57	0.54	0.71	0.40	0.76	0.56	0.53	0.64	0.62	0.47
France	0.90	53	0.66	0.65	0.68	0.49	0.78	0.79	0.80	0.82	0.80	0.77
Georgia	0.97	76	0.64	0.80	0.90	0.69	0.93	0.95	0.95	0.94	0.91	0.95
Germany	0.89	50	0.62	0.67	0.73	0.60	0.73	0.76	0.66	0.78	0.82	0.63
Hong Kong SAR	0.80	37	0.50	0.57	0.65	0.59	0.73	0.68	0.73	0.61	0.67	0.21
Hungary	0.90	53	0.72	0.69	0.79	0.71	0.77	0.81	0.55	0.72	0.80	0.67
Iran, Islamic Rep. of	0.91	57	0.66	0.60	0.72	0.76	0.85	0.82	0.83	0.82	0.78	0.69
Ireland	0.85	46	0.59	0.60	0.70	0.53	0.70	0.68	0.61	0.74	0.82	0.73
Israel	0.96	72	0.71	0.82	0.76	0.89	0.80	0.94	0.89	0.86	0.89	0.92
Italy	0.94	66	0.61	0.75	0.64	0.73	0.69	0.92	0.92	0.92	0.91	0.91
Kazakhstan	0.95	70	0.79	0.80	0.80	0.58	0.83	0.93	0.92	0.88	0.89	0.88
Kuwait	0.91	58	0.49	0.58	0.71	0.79	0.84	0.86	0.84	0.81	0.87	0.76
Latvia	0.83	42	0.63	0.59	0.82	0.37	0.68	0.74	0.64	0.64	0.69	0.56
Lithuania	0.79	36	0.52	0.48	0.69	0.58	0.70	0.60	0.61	0.64	0.61	0.49
Macao SAR	0.73	31	0.72	0.47	0.65	0.28	0.70	0.28	0.76	0.45	0.72	0.03
Malta	0.93	61	0.56	0.77	0.61	0.82	0.83	0.85	0.85	0.86	0.80	0.82
Morocco	0.95	68	0.63	0.74	0.84	0.86	0.90	0.90	0.88	0.86	0.84	0.75
Netherlands	0.85	43	0.39	0.46	0.78	0.59	0.82	0.75	0.54	0.80	0.79	0.43
New Zealand	0.88	49	0.64	0.66	0.63	0.52	0.77	0.75	0.75	0.75	0.79	0.74
Northern Ireland	0.74	34	0.30	0.47	0.76	0.55	0.77	0.53	0.53	0.63	0.62	0.48
Norway (5)	0.85	45	0.60	0.67	0.68	0.65	0.72	0.69	0.49	0.70	0.68	0.77
Oman	0.96	74	0.73	0.81	0.83	0.93	0.89	0.91	0.92	0.90	0.76	0.89
Poland	0.81	38	0.50	0.54	0.57	0.63	0.73	0.65	0.57	0.65	0.71	0.55
Portugal	0.89	53	0.49	0.67	0.77	0.67	0.75	0.79	0.76	0.77	0.77	0.81
Qatar	0.92	61	0.62	0.68	0.73	0.85	0.82	0.82	0.79	0.81	0.80	0.85
Russian Federation	0.77	33	0.60	0.62	0.64	0.60	0.58	0.53	0.60	0.56	0.59	0.35
Saudi Arabia	0.95	70	0.53	0.73	0.81	0.88	0.90	0.90	0.91	0.91	0.88	0.84
Singapore	0.89	50	0.75	0.72	0.76	0.64	0.78	0.67	0.72	0.71	0.82	0.46
Slovak Republic	0.89	51	0.65	0.74	0.70	0.67	0.76	0.80	0.67	0.65	0.75	0.74
Slovenia	0.90	54	0.61	0.67	0.73	0.76	0.77	0.83	0.80	0.76	0.76	0.64
South Africa	0.90	51	0.67	0.72	0.69	0.67	0.72	0.79	0.78	0.72	0.75	0.65
Spain	0.94	69	0.67	0.84	0.77	0.80	0.76	0.91	0.92	0.87	0.86	0.89
Sweden	0.85	44	0.50	0.63	0.76	0.47	0.72	0.78	0.55	0.72	0.76	0.68
Trinidad and Tobago	0.90	54	0.58	0.68	0.58	0.67	0.82	0.85	0.85	0.79	0.83	0.62
United Arab Emirates	0.93	64	0.66	0.68	0.69	0.86	0.84	0.89	0.86	0.82	0.81	0.87
United States	0.87	47	0.59	0.65	0.69	0.61	0.78	0.70	0.68	0.71	0.78	0.66
Benchmarking Participants												
Buenos Aires, Argentina	0.86	49	0.46	0.58	0.65	0.69	0.75	0.80	0.79	0.73	0.77	0.69
Ontario, Canada	0.91	57	0.69	0.68	0.74	0.71	0.79	0.79	0.78	0.81	0.81	0.72
Quebec, Canada	0.83	43	0.58	0.58	0.73	0.35	0.74	0.74	0.72	0.69	0.74	0.58
Denmark (3)	0.84	43	0.60	0.55	0.68	0.51	0.78	0.67	0.49	0.77	0.75	0.65
Norway (4)	0.87	47	0.54	0.46	0.77	0.49	0.79	0.78	0.69	0.73	0.70	0.80
Moscow City, Russian Fed.	0.75	31	0.57	0.56	0.56	0.58	0.67	0.67	0.57	0.47	0.54	0.20
Eng/Afr/Zulu - RSA (5)	0.91	56	0.69	0.72	0.74	0.73	0.77	0.80	0.82	0.78	0.75	0.69
Andalusia, Spain	0.95	70	0.72	0.85	0.70	0.80	0.79	0.93	0.92	0.87	0.84	0.90
Madrid, Spain	0.88	55	0.48	0.51	0.71	0.80	0.74	0.84	0.81	0.77	0.81	0.82
Abu Dhabi, UAE	0.93	63	0.67	0.68	0.71	0.82	0.79	0.88	0.86	0.86	0.79	0.88
Dubai, UAE	0.94	67	0.56	0.73	0.72	0.89	0.87	0.89	0.87	0.85	0.82	0.89

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *School Discipline* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.24	0.06	0.05
Austria	0.10	0.01	0.02
Azerbaijan	-0.06	0.00	0.01
Bahrain	0.18	0.03	0.02
Belgium (Flemish)	0.17	0.03	0.02
Belgium (French)	0.12	0.01	0.02
Bulgaria	0.13	0.02	0.02
Canada	0.16	0.03	0.02
Chile	0.20	0.04	0.05
Chinese Taipei	0.03	0.00	0.00
Czech Republic	0.04	0.00	0.00
Denmark	0.08	0.01	0.01
Egypt	0.13	0.02	0.01
England	0.15	0.02	0.01
Finland	0.06	0.00	0.00
France	0.10	0.01	0.01
Georgia	0.07	0.01	0.00
Germany	0.26	0.07	0.09
Hong Kong SAR	0.17	0.03	0.01
Hungary	0.25	0.06	0.04
Iran, Islamic Rep. of	0.22	0.05	0.05
Ireland	0.12	0.01	0.03
Israel	0.24	0.06	0.07
Italy	0.09	0.01	0.00
Kazakhstan	0.01	0.00	0.01
Kuwait	0.12	0.02	0.02
Latvia	0.08	0.01	0.01
Lithuania	0.04	0.00	0.00
Macao SAR	0.02	0.00	0.01
Malta	0.09	0.01	0.01
Morocco	0.04	0.00	0.00
Netherlands	0.12	0.01	0.02
New Zealand	0.23	0.05	0.06
Northern Ireland	0.14	0.02	0.00
Norway (5)	0.08	0.01	0.01
Oman	0.09	0.01	0.01
Poland	0.07	0.01	0.00
Portugal	0.17	0.03	0.02
Qatar	0.06	0.00	0.01
Russian Federation	0.03	0.00	0.00
Saudi Arabia	0.36	0.13	0.10
Singapore	0.12	0.01	0.01
Slovak Republic	0.16	0.03	0.08
Slovenia	0.01	0.00	0.00
South Africa	0.17	0.03	0.02
Spain	0.16	0.02	0.02
Sweden	0.13	0.02	0.02
Trinidad and Tobago	0.10	0.01	0.01
United Arab Emirates	0.24	0.06	0.03
United States	0.22	0.05	0.03
International Median	0.12	0.01	0.01
Benchmarking Participants			
Buenos Aires, Argentina	0.11	0.01	0.02
Ontario, Canada	0.16	0.03	0.03
Quebec, Canada	0.15	0.02	0.01
Denmark (3)	0.05	0.00	0.01
Norway (4)	0.06	0.00	0.00
Moscow City, Russian Fed.	-0.04	0.00	0.00
Eng/Afr/Zulu - RSA (5)	0.12	0.01	0.01
Andalusia, Spain	0.08	0.01	0.01
Madrid, Spain	0.11	0.01	0.02
Abu Dhabi, UAE	0.14	0.02	0.01
Dubai, UAE	0.26	0.07	0.04

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

School Emphasis on Academic Success – Principals' Reports Scale

The *School Emphasis on Academic Success – Principals' Reports* (EAS) scale was created based on principals' responses characterizing the 12 aspects of school climate described below.

Items in the PIRLS 2016 *School Emphasis on Academic Success – Principals' Reports* Scale[1]

How would you characterize each of the following within your school?

		Very high	High	Medium	Low	Very low
ACBG13A	1) Teachers' understanding of the school's curricular goals	◯	◯	◯	◯	◯
ACBG13B	2) Teachers' degree of success in implementing the school's curriculum	◯	◯	◯	◯	◯
ACBG13C	3) Teachers' expectations for student achievement	◯	◯	◯	◯	◯
ACBG13D	4) Teachers' ability to inspire students	◯	◯	◯	◯	◯
ACBG13E	5) Collaboration between school leadership (including master teachers) and teachers to plan instruction	◯	◯	◯	◯	◯
ACBG13F	6) Parental involvement in school activities	◯	◯	◯	◯	◯
ACBG13G	7) Parental commitment to ensure that students are ready to learn	◯	◯	◯	◯	◯
ACBG13H	8) Parental expectations for student achievement	◯	◯	◯	◯	◯
ACBG13I	9) Parental support for student achievement	◯	◯	◯	◯	◯
ACBG13J	10) Students' desire to do well in school	◯	◯	◯	◯	◯
ACBG13K	11) Students' ability to reach school's academic goals	◯	◯	◯	◯	◯
ACBG13L	12) Students' respect for classmates who excel academically	◯	◯	◯	◯	◯

Very High Emphasis | 12.9 | High Emphasis | 9.2 | Medium Emphasis

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Low" and "Very low" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the PIRLS 2016 *School Emphasis on Academic Success - Principals' Reports* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ACBG13A	-1.31219	-3.35435	-0.01880	3.37315	1.07
ACBG13B	-0.96099	-3.90556	0.19733	3.70823	0.96
ACBG13C	-0.82391	-3.49024	0.00835	3.48189	0.89
ACBG13D	-0.59035	-3.43001	0.05756	3.37245	0.99
ACBG13E	-0.68959	-2.63930	-0.22702	2.86632	1.20
ACBG13F	1.15412	-2.48630	0.08282	2.40348	1.17
ACBG13G	1.42236	-2.80818	0.13531	2.67287	0.85
ACBG13H	0.00207	-2.69019	-0.24463	2.93482	1.09
ACBG13I	1.29200	-2.95076	0.10280	2.84796	0.91
ACBG13J	0.16604	-3.49673	0.03249	3.46424	0.87
ACBG13K	0.50539	-4.02618	0.21215	3.81403	0.87
ACBG13L	-0.16495	-3.14717	-0.20095	3.34812	1.14

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *School Emphasis on Academic Success - Principals' Reports* Scale

Scale Transformation Constants	
A = 9.088617	Transformed Scale Score = 9.088617 + 1.147876 • Logit Scale Score
B = 1.147876	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *School Emphasis on Academic Success - Principals' Reports* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	1.06095	
1	2.49562	
2	3.26563	
3	3.84274	
4	4.33403	
5	4.77520	
6	5.18627	
7	5.57770	
8	5.95426	
9	6.31720	
10	6.66619	
11	7.00120	
12	7.32360	
13	7.63552	
14	7.93995	
15	8.24014	
16	8.53923	
17	8.83993	
18	9.14415	9.2
19	9.45276	
20	9.76546	
21	10.08096	
22	10.39737	
23	10.71295	
24	11.02647	
25	11.33774	
26	11.64774	
27	11.95831	
28	12.27223	
29	12.59338	
30	12.92713	12.9
31	13.28181	
32	13.66669	
33	14.10420	
34	14.63433	
35	15.35698	
36	16.74931	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016
***School Emphasis on Academic Success - Principals' Reports* Scale**

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			ACBG13A	ACBG13B	ACBG13C	ACBG13D	ACBG13E	ACBG13F	ACBG13G	ACBG13H	ACBG13I	ACBG13J	ACBG13K	ACBG13L
Australia	0.93	56	0.75	0.78	0.78	0.68	0.63	0.61	0.81	0.77	0.82	0.80	0.80	0.72
Austria	0.83	37	0.43	0.60	0.73	0.34	0.27	0.66	0.82	0.63	0.78	0.62	0.71	0.41
Azerbaijan	0.89	46	0.74	0.75	0.68	0.72	0.72	0.68	0.72	0.62	0.64	0.62	0.61	0.57
Bahrain	0.91	51	0.67	0.76	0.79	0.81	0.64	0.66	0.68	0.64	0.73	0.77	0.81	0.58
Belgium (Flemish)	0.83	36	0.41	0.52	0.47	0.41	0.34	0.67	0.82	0.65	0.81	0.58	0.68	0.64
Belgium (French)	0.84	38	0.52	0.61	0.49	0.55	0.41	0.66	0.77	0.65	0.77	0.68	0.70	0.53
Bulgaria	0.90	50	0.49	0.73	0.71	0.69	0.54	0.78	0.78	0.63	0.84	0.75	0.80	0.65
Canada	0.88	45	0.54	0.67	0.69	0.64	0.50	0.68	0.73	0.72	0.77	0.71	0.70	0.60
Chile	0.92	54	0.71	0.72	0.80	0.76	0.72	0.73	0.74	0.78	0.75	0.78	0.70	0.65
Chinese Taipei	0.89	46	0.58	0.52	0.74	0.73	0.57	0.61	0.72	0.68	0.68	0.80	0.81	0.64
Czech Republic	0.85	38	0.54	0.58	0.59	0.61	0.60	0.53	0.71	0.57	0.63	0.74	0.73	0.52
Denmark	0.91	50	0.73	0.68	0.75	0.61	0.54	0.55	0.82	0.72	0.71	0.83	0.74	0.75
Egypt	0.91	50	0.66	0.66	0.68	0.59	0.65	0.70	0.77	0.75	0.74	0.74	0.79	0.70
England	0.90	47	0.64	0.72	0.65	0.71	0.65	0.66	0.77	0.75	0.74	0.67	0.72	0.51
Finland	0.89	46	0.58	0.61	0.73	0.72	0.57	0.58	0.77	0.74	0.71	0.71	0.71	0.63
France	0.85	39	0.40	0.61	0.56	0.58	0.48	0.63	0.70	0.66	0.72	0.77	0.74	0.59
Georgia	0.87	41	0.68	0.69	0.66	0.66	0.61	0.61	0.68	0.65	0.71	0.61	0.53	0.54
Germany	0.84	38	0.52	0.60	0.66	0.49	0.40	0.63	0.78	0.58	0.73	0.64	0.70	0.55
Hong Kong SAR	0.87	42	0.64	0.66	0.62	0.65	0.62	0.65	0.67	0.61	0.68	0.71	0.73	0.47
Hungary	0.88	45	0.43	0.68	0.61	0.71	0.39	0.69	0.81	0.72	0.81	0.72	0.68	0.66
Iran, Islamic Rep. of	0.91	50	0.67	0.71	0.65	0.70	0.58	0.73	0.81	0.68	0.79	0.78	0.69	0.67
Ireland	0.89	47	0.58	0.62	0.75	0.63	0.40	0.58	0.82	0.81	0.81	0.81	0.72	0.54
Israel	0.86	41	0.71	0.71	0.69	0.67	0.66	0.50	0.66	0.60	0.67	0.69	0.60	0.46
Italy	0.86	40	0.63	0.69	0.55	0.77	0.70	0.51	0.64	0.52	0.62	0.71	0.59	0.56
Kazakhstan	0.90	49	0.56	0.66	0.67	0.74	0.69	0.67	0.79	0.64	0.77	0.71	0.77	0.66
Kuwait	0.92	54	0.71	0.74	0.78	0.77	0.61	0.66	0.76	0.70	0.65	0.82	0.81	0.74
Latvia	0.86	40	0.53	0.68	0.57	0.65	0.55	0.76	0.70	0.48	0.71	0.71	0.70	0.43
Lithuania	0.89	45	0.64	0.67	0.68	0.75	0.68	0.67	0.73	0.53	0.70	0.67	0.67	0.62
Macao SAR	0.79	32	0.51	0.61	0.60	0.71	0.40	0.49	0.70	0.55	0.52	0.71	0.50	0.41
Malta	0.88	44	0.63	0.74	0.71	0.65	0.59	0.61	0.71	0.73	0.73	0.66	0.67	0.47
Morocco	0.90	48	0.68	0.73	0.75	0.69	0.70	0.67	0.76	0.63	0.72	0.65	0.71	0.62
Netherlands	0.85	38	0.53	0.62	0.72	0.77	0.58	0.43	0.58	0.57	0.54	0.71	0.62	0.63
New Zealand	0.91	50	0.71	0.75	0.78	0.68	0.64	0.67	0.72	0.67	0.76	0.70	0.76	0.63
Northern Ireland	0.90	49	0.59	0.73	0.75	0.73	0.74	0.52	0.66	0.65	0.71	0.78	0.78	0.67
Norway (5)	0.90	49	0.67	0.79	0.74	0.71	0.63	0.72	0.76	0.66	0.74	0.58	0.74	0.62
Oman	0.88	44	0.62	0.66	0.69	0.71	0.63	0.61	0.76	0.64	0.69	0.69	0.68	0.59
Poland	0.87	42	0.60	0.66	0.47	0.67	0.57	0.65	0.77	0.70	0.71	0.67	0.66	0.57
Portugal	0.89	46	0.55	0.68	0.69	0.63	0.59	0.73	0.71	0.72	0.77	0.75	0.69	0.59
Qatar	0.89	47	0.64	0.69	0.75	0.74	0.60	0.68	0.79	0.63	0.71	0.62	0.71	0.61
Russian Federation	0.87	41	0.58	0.73	0.70	0.70	0.56	0.63	0.64	0.47	0.74	0.65	0.72	0.55
Saudi Arabia	0.92	53	0.68	0.74	0.68	0.75	0.74	0.68	0.75	0.71	0.76	0.80	0.75	0.71
Singapore	0.93	56	0.67	0.74	0.79	0.74	0.61	0.69	0.84	0.80	0.78	0.81	0.79	0.66
Slovak Republic	0.86	41	0.52	0.67	0.63	0.61	0.64	0.64	0.71	0.58	0.67	0.71	0.72	0.52
Slovenia	0.85	38	0.60	0.64	0.54	0.73	0.66	0.57	0.66	0.36	0.64	0.63	0.68	0.61
South Africa	0.90	49	0.62	0.58	0.67	0.65	0.69	0.70	0.75	0.64	0.77	0.77	0.80	0.69
Spain	0.90	48	0.61	0.74	0.77	0.64	0.57	0.67	0.80	0.78	0.79	0.67	0.66	0.54
Sweden	0.91	52	0.58	0.67	0.76	0.74	0.60	0.69	0.79	0.83	0.77	0.81	0.70	0.65
Trinidad and Tobago	0.90	49	0.57	0.63	0.67	0.65	0.58	0.77	0.79	0.64	0.76	0.79	0.79	0.68
United Arab Emirates	0.92	53	0.72	0.76	0.78	0.75	0.67	0.67	0.74	0.69	0.76	0.74	0.76	0.62
United States	0.93	56	0.67	0.73	0.72	0.77	0.59	0.76	0.82	0.81	0.81	0.81	0.79	0.69
Benchmarking Participants														
Buenos Aires, Argentina	0.90	50	0.69	0.78	0.66	0.63	0.58	0.71	0.81	0.73	0.80	0.76	0.76	0.50
Ontario, Canada	0.88	45	0.54	0.68	0.74	0.67	0.42	0.63	0.69	0.77	0.76	0.77	0.71	0.58
Quebec, Canada	0.88	43	0.68	0.69	0.58	0.70	0.64	0.73	0.70	0.65	0.75	0.55	0.61	0.55
Denmark (3)	0.90	50	0.73	0.67	0.75	0.61	0.54	0.56	0.82	0.72	0.73	0.83	0.73	0.73
Norway (4)	0.90	49	0.66	0.79	0.74	0.73	0.63	0.73	0.76	0.66	0.74	0.58	0.74	0.62
Moscow City, Russian Fed.	0.84	37	0.66	0.76	0.62	0.73	0.66	0.49	0.63	0.25	0.65	0.57	0.63	0.44
Eng/Afr/Zulu - RSA (5)	0.90	48	0.46	0.56	0.69	0.71	0.67	0.70	0.80	0.61	0.80	0.79	0.76	0.71
Andalusia, Spain	0.88	44	0.58	0.75	0.74	0.61	0.60	0.72	0.75	0.77	0.75	0.62	0.61	0.39
Madrid, Spain	0.91	52	0.61	0.69	0.76	0.67	0.61	0.73	0.83	0.84	0.83	0.71	0.75	0.62
Abu Dhabi, UAE	0.89	46	0.74	0.76	0.80	0.77	0.61	0.54	0.66	0.56	0.64	0.72	0.69	0.59
Dubai, UAE	0.93	57	0.73	0.76	0.83	0.78	0.72	0.74	0.72	0.74	0.77	0.76	0.79	0.69

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *School Emphasis on Academic Success - Principals' Reports* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.24	0.06	0.06
Austria	0.22	0.05	0.04
Azerbaijan	-0.04	0.00	0.02
Bahrain	0.29	0.08	0.06
Belgium (Flemish)	0.15	0.02	0.02
Belgium (French)	0.26	0.07	0.05
Bulgaria	0.35	0.12	0.08
Canada	0.20	0.04	0.04
Chile	0.22	0.05	0.03
Chinese Taipei	0.13	0.02	0.01
Czech Republic	0.12	0.01	0.00
Denmark	0.13	0.02	0.02
Egypt	0.22	0.05	0.03
England	0.18	0.03	0.02
Finland	0.09	0.01	0.01
France	0.05	0.00	0.00
Georgia	0.07	0.01	0.00
Germany	0.31	0.10	0.08
Hong Kong SAR	0.15	0.02	0.00
Hungary	0.29	0.09	0.04
Iran, Islamic Rep. of	0.21	0.04	0.03
Ireland	0.17	0.03	0.04
Israel	0.12	0.01	0.02
Italy	0.02	0.00	0.01
Kazakhstan	0.09	0.01	0.01
Kuwait	0.23	0.05	0.04
Latvia	0.16	0.03	0.02
Lithuania	0.26	0.07	0.07
Macao SAR	0.16	0.02	0.02
Malta	0.06	0.00	0.00
Morocco	0.34	0.11	0.12
Netherlands	0.12	0.01	0.01
New Zealand	0.18	0.03	0.03
Northern Ireland	0.10	0.01	0.00
Norway (5)	0.16	0.02	0.02
Oman	0.25	0.06	0.05
Poland	0.14	0.02	0.01
Portugal	0.26	0.07	0.05
Qatar	0.15	0.02	0.03
Russian Federation	0.18	0.03	0.04
Saudi Arabia	0.26	0.07	0.05
Singapore	0.19	0.04	0.03
Slovak Republic	0.30	0.09	0.04
Slovenia	0.08	0.01	0.00
South Africa	0.11	0.01	0.03
Spain	0.17	0.03	0.02
Sweden	0.20	0.04	0.03
Trinidad and Tobago	0.27	0.07	0.05
United Arab Emirates	0.30	0.09	0.08
United States	0.19	0.03	0.03
International Median	0.18	0.03	0.03
Benchmarking Participants			
Buenos Aires, Argentina	0.23	0.05	0.06
Ontario, Canada	0.17	0.03	0.03
Quebec, Canada	0.17	0.03	0.02
Denmark (3)	0.08	0.01	0.01
Norway (4)	0.11	0.01	0.01
Moscow City, Russian Fed.	0.09	0.01	0.02
Eng/Afr/Zulu - RSA (5)	0.19	0.04	0.03
Andalusia, Spain	0.20	0.04	0.02
Madrid, Spain	0.21	0.05	0.05
Abu Dhabi, UAE	0.20	0.04	0.04
Dubai, UAE	0.38	0.14	0.11

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

School Emphasis on Academic Success – Teachers' Reports Scale

The *School Emphasis on Academic Success – Teachers' Reports* (EAS) scale was created based on teachers' responses characterizing the 12 aspects of school climate described below.

Items in the PIRLS 2016 *School Emphasis on Academic Success – Teachers' Reports* Scale[1]

How would you characterize each of the following within your school?

		Very high	High	Medium	Low	Very low
ATBG07A	1) Teachers' understanding of the school's curricular goals	◯	◯	◯	◯	◯
ATBG07B	2) Teachers' degree of success in implementing the school's curriculum	◯	◯	◯	◯	◯
ATBG07C	3) Teachers' expectations for student achievement	◯	◯	◯	◯	◯
ATBG07D	4) Teachers' ability to inspire students	◯	◯	◯	◯	◯
ATBG07E	5) Collaboration between school leadership (including master teachers) and teachers to plan instruction	◯	◯	◯	◯	◯
ATBG07F	6) Parental involvement in school activities	◯	◯	◯	◯	◯
ATBG07G	7) Parental commitment to ensure that students are ready to learn	◯	◯	◯	◯	◯
ATBG07H	8) Parental expectations for student achievement	◯	◯	◯	◯	◯
ATBG07I	9) Parental support for student achievement	◯	◯	◯	◯	◯
ATBG07J	10) Students' desire to do well in school	◯	◯	◯	◯	◯
ATBG07K	11) Students' ability to reach school's academic goals	◯	◯	◯	◯	◯
ATBG07L	12) Students' respect for classmates who excel academically	◯	◯	◯	◯	◯

Very High Emphasis | 12.8 | High Emphasis | 9.2 | Medium Emphasis

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Low" and "Very low" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the PIRLS 2016 *School Emphasis on Academic Success - Teachers' Reports* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ATBG07A	-1.32583	-2.86202	-0.08581	2.94783	1.08
ATBG07B	-0.93336	-3.25924	0.00909	3.25015	0.97
ATBG07C	-0.90529	-3.15010	-0.01642	3.16652	1.05
ATBG07D	-0.95737	-3.07196	-0.05710	3.12906	1.00
ATBG07E	-0.12321	-1.65307	-0.24929	1.90236	1.33
ATBG07F	0.99236	-2.17211	0.07001	2.10210	1.04
ATBG07G	1.25870	-2.51430	0.14423	2.37007	0.85
ATBG07H	0.10750	-2.38394	-0.24263	2.62657	1.05
ATBG07I	1.12745	-2.70509	0.13541	2.56968	0.86
ATBG07J	0.20474	-2.87243	-0.04754	2.91997	0.91
ATBG07K	0.68915	-3.39428	0.15216	3.24212	0.85
ATBG07L	-0.13484	-2.38835	-0.40389	2.79224	1.12

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *School Emphasis on Academic Success - Teachers' Reports* Scale

Scale Transformation Constants	
A = 9.085861	Transformed Scale Score = 9.085861 + 1.287929 • Logit Scale Score
B = 1.287929	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *School Emphasis on Academic Success - Teachers' Reports* Scale

Raw Score	Transformed Scale Score	Cutpoint
1	2.31036	
2	3.17847	
3	3.83356	
4	4.38451	
5	4.87013	
6	5.30628	
7	5.70360	
8	6.06975	
9	6.41275	
10	6.73719	
11	7.04747	
12	7.34668	
13	7.64101	
14	7.93095	
15	8.22006	
16	8.51056	
17	8.80401	
18	9.10134	9.2
19	9.40263	
20	9.70726	
21	10.01421	
22	10.32239	
23	10.63104	
24	10.93982	
25	11.24966	
26	11.56193	
27	11.87884	
28	12.20341	
29	12.53973	
30	12.89347	12.8
31	13.27297	
32	13.69227	
33	14.16929	
34	14.75380	
35	15.55747	
36	17.11436	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *School Emphasis on Academic Success - Teachers' Reports* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item											
			ATBG07A	ATBG07B	ATBG07C	ATBG07D	ATBG07E	ATBG07F	ATBG07G	ATBG07H	ATBG07I	ATBG07J	ATBG07K	ATBG07L
Australia	0.90	48	0.56	0.63	0.70	0.56	0.44	0.69	0.84	0.79	0.80	0.77	0.76	0.67
Austria	0.82	36	0.40	0.59	0.55	0.48	0.45	0.72	0.81	0.56	0.76	0.63	0.72	0.34
Azerbaijan	0.86	39	0.58	0.66	0.63	0.60	0.66	0.71	0.70	0.63	0.66	0.48	0.66	0.50
Bahrain	0.91	52	0.56	0.68	0.73	0.71	0.61	0.76	0.80	0.72	0.82	0.73	0.79	0.71
Belgium (Flemish)	0.79	31	0.36	0.47	0.59	0.48	0.55	0.50	0.74	0.51	0.74	0.56	0.65	0.42
Belgium (French)	0.82	36	0.45	0.50	0.48	0.49	0.39	0.63	0.75	0.60	0.75	0.70	0.74	0.56
Bulgaria	0.91	49	0.52	0.65	0.73	0.65	0.63	0.74	0.78	0.70	0.78	0.76	0.78	0.60
Canada	0.88	45	0.54	0.60	0.65	0.63	0.48	0.68	0.80	0.72	0.80	0.73	0.77	0.53
Chile	0.77	32	-0.20	-0.21	-0.24	-0.18	-0.18	0.66	0.84	0.74	0.78	0.70	0.69	0.59
Chinese Taipei	0.83	35	0.51	0.42	0.48	0.52	0.48	0.67	0.77	0.65	0.72	0.69	0.63	0.45
Czech Republic	0.86	41	0.58	0.70	0.65	0.61	0.57	0.63	0.66	0.57	0.67	0.68	0.74	0.56
Denmark	0.87	41	0.60	0.55	0.66	0.52	0.51	0.67	0.77	0.56	0.77	0.72	0.70	0.60
Egypt	0.90	48	0.62	0.67	0.63	0.57	0.62	0.74	0.82	0.74	0.79	0.74	0.78	0.57
England	0.89	47	0.62	0.63	0.62	0.59	0.55	0.68	0.82	0.75	0.77	0.73	0.78	0.63
Finland	0.82	35	0.39	0.42	0.63	0.56	0.40	0.64	0.67	0.70	0.73	0.69	0.64	0.50
France	0.83	38	0.34	0.59	0.47	0.42	0.37	0.68	0.83	0.64	0.80	0.73	0.76	0.46
Georgia	0.88	43	0.66	0.70	0.62	0.55	0.63	0.69	0.73	0.61	0.81	0.69	0.58	0.57
Germany	0.83	36	0.41	0.53	0.57	0.54	0.43	0.61	0.79	0.62	0.79	0.61	0.73	0.47
Hong Kong SAR	0.87	42	0.52	0.56	0.51	0.63	0.54	0.70	0.71	0.68	0.74	0.78	0.77	0.57
Hungary	0.91	50	0.57	0.69	0.67	0.64	0.42	0.75	0.83	0.76	0.80	0.76	0.77	0.72
Iran, Islamic Rep. of	0.91	49	0.72	0.73	0.60	0.73	0.71	0.78	0.78	0.63	0.76	0.70	0.71	0.54
Ireland	0.90	48	0.47	0.64	0.74	0.60	0.49	0.69	0.84	0.78	0.85	0.80	0.74	0.55
Israel	0.89	47	0.70	0.74	0.65	0.61	0.64	0.55	0.67	0.68	0.74	0.78	0.76	0.67
Italy	0.88	44	0.69	0.66	0.71	0.72	0.69	0.64	0.78	0.45	0.73	0.64	0.69	0.49
Kazakhstan	0.90	47	0.59	0.70	0.67	0.69	0.66	0.60	0.70	0.68	0.74	0.69	0.79	0.72
Kuwait	0.91	51	0.64	0.66	0.68	0.74	0.66	0.69	0.83	0.63	0.77	0.77	0.81	0.69
Latvia	0.86	40	0.67	0.71	0.55	0.67	0.72	0.64	0.66	0.45	0.70	0.58	0.65	0.46
Lithuania	0.83	36	0.48	0.69	0.68	0.62	0.49	0.55	0.69	0.73	0.64	0.55	0.61	0.40
Macao SAR	0.90	49	0.62	0.73	0.67	0.76	0.74	0.70	0.71	0.70	0.70	0.71	0.75	0.63
Malta	0.85	40	0.61	0.64	0.64	0.59	0.44	0.42	0.66	0.73	0.73	0.73	0.71	0.58
Morocco	0.91	50	0.58	0.66	0.67	0.70	0.70	0.72	0.79	0.70	0.74	0.74	0.81	0.65
Netherlands	0.78	30	0.32	0.38	0.54	0.51	0.39	0.52	0.69	0.62	0.75	0.61	0.63	0.47
New Zealand	0.89	47	0.61	0.64	0.65	0.56	0.56	0.73	0.77	0.70	0.79	0.73	0.75	0.68
Northern Ireland	0.88	44	0.71	0.74	0.68	0.57	0.61	0.64	0.67	0.54	0.67	0.78	0.76	0.54
Norway (5)	0.87	42	0.49	0.67	0.60	0.60	0.56	0.68	0.78	0.63	0.75	0.73	0.70	0.55
Oman	0.91	49	0.61	0.64	0.73	0.62	0.62	0.75	0.79	0.75	0.75	0.72	0.73	0.69
Poland	0.89	47	0.47	0.62	0.65	0.59	0.65	0.73	0.80	0.64	0.81	0.73	0.75	0.67
Portugal	0.90	48	0.64	0.71	0.72	0.64	0.57	0.75	0.80	0.76	0.80	0.70	0.71	0.48
Qatar	0.89	47	0.66	0.70	0.71	0.74	0.60	0.68	0.73	0.58	0.69	0.70	0.77	0.62
Russian Federation	0.86	41	0.66	0.69	0.60	0.65	0.49	0.70	0.73	0.48	0.71	0.68	0.69	0.51
Saudi Arabia	0.91	51	0.69	0.69	0.51	0.76	0.61	0.73	0.79	0.74	0.81	0.72	0.77	0.69
Singapore	0.90	48	0.62	0.64	0.63	0.67	0.60	0.70	0.82	0.66	0.77	0.76	0.78	0.62
Slovak Republic	0.88	44	0.56	0.64	0.65	0.59	0.54	0.69	0.74	0.59	0.75	0.75	0.78	0.62
Slovenia	0.78	30	0.57	0.62	0.54	0.64	0.66	0.55	0.60	0.36	0.52	0.38	0.57	0.46
South Africa	0.92	50	0.53	0.63	0.60	0.59	0.54	0.76	0.80	0.74	0.84	0.81	0.83	0.74
Spain	0.87	42	0.54	0.66	0.64	0.53	0.50	0.71	0.82	0.76	0.76	0.63	0.67	0.51
Sweden	0.85	40	0.46	0.51	0.63	0.62	0.50	0.71	0.77	0.72	0.75	0.74	0.62	0.44
Trinidad and Tobago	0.91	52	0.53	0.63	0.64	0.69	0.67	0.79	0.85	0.72	0.84	0.80	0.76	0.64
United Arab Emirates	0.91	52	0.67	0.70	0.69	0.71	0.66	0.72	0.80	0.67	0.81	0.77	0.81	0.64
United States	0.89	47	0.41	0.58	0.55	0.59	0.45	0.78	0.86	0.82	0.84	0.72	0.73	0.69
Benchmarking Participants														
Buenos Aires, Argentina	0.89	46	0.46	0.65	0.60	0.59	0.58	0.74	0.82	0.74	0.79	0.73	0.80	0.52
Ontario, Canada	0.90	50	0.57	0.67	0.67	0.69	0.46	0.73	0.82	0.78	0.84	0.75	0.84	0.59
Quebec, Canada	0.82	35	0.48	0.49	0.66	0.47	0.39	0.59	0.77	0.69	0.77	0.65	0.68	0.28
Denmark (3)	0.88	44	0.55	0.62	0.65	0.52	0.50	0.64	0.75	0.74	0.78	0.79	0.75	0.59
Norway (4)	0.88	44	0.63	0.73	0.60	0.68	0.56	0.60	0.70	0.69	0.77	0.66	0.70	0.61
Moscow City, Russian Fed.	0.81	34	0.57	0.55	0.67	0.49	0.59	0.64	0.61	0.45	0.64	0.61	0.71	0.36
Eng/Afr/Zulu - RSA (5)	0.92	52	0.60	0.64	0.56	0.69	0.64	0.85	0.85	0.51	0.86	0.84	0.76	0.75
Andalusia, Spain	0.90	48	0.55	0.77	0.71	0.54	0.54	0.71	0.86	0.79	0.83	0.69	0.71	0.51
Madrid, Spain	0.87	45	0.55	0.61	0.76	0.51	0.29	0.78	0.82	0.77	0.82	0.70	0.73	0.53
Abu Dhabi, UAE	0.91	53	0.73	0.74	0.72	0.71	0.64	0.66	0.79	0.68	0.85	0.80	0.80	0.57
Dubai, UAE	0.94	59	0.73	0.75	0.77	0.75	0.68	0.81	0.81	0.67	0.83	0.81	0.84	0.71

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *School Emphasis on Academic Success - Teachers' Reports* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.25	0.06	0.05
Austria	0.22	0.05	0.04
Azerbaijan	0.12	0.01	0.01
Bahrain	0.25	0.06	0.05
Belgium (Flemish)	0.16	0.03	0.01
Belgium (French)	0.24	0.06	0.05
Bulgaria	0.34	0.11	0.08
Canada	0.12	0.01	0.01
Chile	-0.01	0.00	0.00
Chinese Taipei	0.05	0.00	0.00
Czech Republic	0.11	0.01	0.01
Denmark	0.11	0.01	0.01
Egypt	0.25	0.06	0.04
England	0.09	0.01	0.01
Finland	0.09	0.01	0.01
France	0.14	0.02	0.02
Georgia	0.13	0.02	0.01
Germany	0.31	0.10	0.08
Hong Kong SAR	0.03	0.00	0.00
Hungary	0.31	0.09	0.06
Iran, Islamic Rep. of	0.11	0.01	0.01
Ireland	0.19	0.03	0.03
Israel	0.12	0.01	0.01
Italy	0.04	0.00	0.00
Kazakhstan	-0.02	0.00	0.00
Kuwait	0.10	0.01	0.01
Latvia	0.13	0.02	0.01
Lithuania	0.19	0.04	0.03
Macao SAR	0.10	0.01	0.02
Malta	0.13	0.02	0.02
Morocco	0.29	0.09	0.09
Netherlands	0.12	0.01	0.01
New Zealand	0.17	0.03	0.03
Northern Ireland	0.11	0.01	0.01
Norway (5)	0.15	0.02	0.02
Oman	0.14	0.02	0.02
Poland	0.06	0.00	0.00
Portugal	0.17	0.03	0.02
Qatar	0.08	0.01	0.01
Russian Federation	0.16	0.03	0.04
Saudi Arabia	0.21	0.04	0.04
Singapore	0.21	0.05	0.04
Slovak Republic	0.27	0.07	0.04
Slovenia	0.04	0.00	0.00
South Africa	0.03	0.00	0.00
Spain	0.13	0.02	0.02
Sweden	0.14	0.02	0.01
Trinidad and Tobago	0.19	0.03	0.04
United Arab Emirates	0.23	0.05	0.05
United States	0.23	0.05	0.06
International Median	0.14	0.02	0.02
Benchmarking Participants			
Buenos Aires, Argentina	0.20	0.04	0.04
Ontario, Canada	0.14	0.02	0.02
Quebec, Canada	0.13	0.02	0.01
Denmark (3)	0.13	0.02	0.01
Norway (4)	0.10	0.01	0.01
Moscow City, Russian Fed.	0.12	0.01	0.02
Eng/Afr/Zulu - RSA (5)	0.05	0.00	0.01
Andalusia, Spain	0.19	0.04	0.03
Madrid, Spain	0.19	0.04	0.03
Abu Dhabi, UAE	0.16	0.02	0.03
Dubai, UAE	0.26	0.07	0.05

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

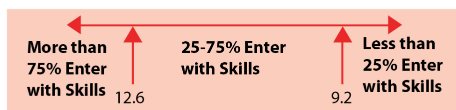
Schools Where Students Enter the Primary Grades with Early Literacy Skills Scale

The *Schools Where Students Enter the Primary Grades with Early Literacy Skills* (ELS) scale was created based on principals' responses about the percentage of children in the school who began first grade with the six key skills described below.

Items in the PIRLS 2016 *Schools Where Students Enter the Primary Grades with Early Literacy Skills* Scale

About how many of the students in your school can do the following when they begin the first grade of primary/elementary school?

			More than 75%	51–75%	25–50%	Less than 25%
T	ACBG16A	1) Recognize most of the letters of the alphabet				
T	ACBG16B	2) Read some words				
T	ACBG16C	3) Read sentences				
	ACBG16D	4) Read a story				
T	ACBG16E	5) Write letters of the alphabet				
T	ACBG16F	6) Write some words				



T Trend item—item was included in the same scale in PIRLS 2011 and was used for linking the PIRLS 2011 and PIRLS 2016 scales.

Item Parameters for the PIRLS 2016 *Schools Where Students Enter the Primary Grades with Early Literacy Skills* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ACBG16A	-1.78233	-1.97973	0.20099	1.77874	1.50
ACBG16B	-0.57140	-2.07581	0.12065	1.95516	0.90
ACBG16C	1.30316	-1.75400	-0.13912	1.89312	0.93
ACBG16D	2.41057	-1.45499	-0.37070	1.82569	1.19
ACBG16E	-1.31646	-2.29289	0.42674	1.86615	1.31
ACBG16F	-0.04354	-1.97416	0.16291	1.81125	1.08

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Schools Where Students Enter the Primary Grades with Early Literacy Skills* Scale

Scale Transformation Constants	
A = 10.926395	Transformed Scale Score = 10.926395 + 0.700871 • Logit Scale Score
B = 0.700871	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Schools Where Students Enter the Primary Grades with Early Literacy Skills* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	7.11411	
1	8.08353	
2	8.66466	
3	9.13106	9.2
4	9.52767	
5	9.87200	
6	10.17733	
7	10.45638	
8	10.71757	
9	10.96860	
10	11.21588	
11	11.46487	
12	11.72126	
13	11.99312	
14	12.29199	
15	12.64157	12.6
16	13.08630	
17	13.70842	
18	14.76593	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Schools Where Students Enter the Primary Grades with Early Literacy Skills* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			ACBG16A	ACBG16B	ACBG16C	ACBG16D	ACBG16E	ACBG16F
Australia	0.94	77	0.87	0.92	0.91	0.77	0.89	0.90
Austria	0.94	78	0.90	0.88	0.94	0.91	0.77	0.88
Azerbaijan	0.95	79	0.84	0.92	0.92	0.83	0.91	0.90
Bahrain	0.94	77	0.82	0.94	0.91	0.73	0.89	0.94
Belgium (Flemish)	0.86	62	0.73	0.83	0.82	0.82	0.69	0.82
Belgium (French)	0.83	53	0.74	0.86	0.68	0.50	0.76	0.78
Bulgaria	0.92	72	0.80	0.91	0.91	0.78	0.82	0.87
Canada	0.93	75	0.83	0.92	0.88	0.77	0.88	0.90
Chile	0.96	84	0.89	0.94	0.93	0.88	0.91	0.93
Chinese Taipei	0.94	78	0.84	0.93	0.92	0.89	0.88	0.85
Czech Republic	0.78	51	0.77	0.73	0.57	0.51	0.85	0.80
Denmark	0.88	64	0.83	0.88	0.82	0.69	0.77	0.79
Egypt	0.93	74	0.89	0.92	0.88	0.63	0.90	0.92
England	0.97	86	0.93	0.95	0.93	0.85	0.94	0.95
Finland	0.86	59	0.67	0.83	0.88	0.74	0.66	0.79
France	0.82	54	0.71	0.79	0.72	0.61	0.75	0.81
Georgia	0.95	80	0.83	0.90	0.94	0.86	0.91	0.94
Germany	0.78	52	0.76	0.83	0.62	0.52	0.76	0.78
Hong Kong SAR	0.95	80	0.88	0.93	0.94	0.91	0.88	0.81
Hungary	0.92	75	0.88	0.92	0.93	0.76	0.80	0.91
Iran, Islamic Rep. of	0.95	81	0.86	0.94	0.94	0.83	0.89	0.93
Ireland	0.87	70	0.78	0.86	0.87	0.65	0.92	0.91
Israel	0.91	70	0.73	0.89	0.88	0.83	0.81	0.88
Italy	0.92	72	0.79	0.90	0.88	0.78	0.85	0.89
Kazakhstan	0.90	68	0.76	0.88	0.90	0.82	0.78	0.81
Kuwait	0.95	81	0.86	0.94	0.94	0.81	0.91	0.93
Latvia	0.90	68	0.68	0.87	0.90	0.82	0.84	0.84
Lithuania	0.89	65	0.71	0.88	0.87	0.76	0.77	0.84
Macao SAR	0.93	76	0.85	0.85	0.92	0.83	0.91	0.86
Malta	0.90	69	0.74	0.90	0.88	0.73	0.82	0.88
Morocco	0.95	82	0.92	0.96	0.91	0.75	0.94	0.94
Netherlands	0.86	61	0.66	0.86	0.81	0.68	0.80	0.84
New Zealand	0.93	79	0.85	0.93	0.92	0.85	0.89	-
Northern Ireland	0.96	86	0.94	0.97	0.96	0.77	0.93	0.96
Norway (5)	0.86	61	0.82	0.87	0.78	0.54	0.81	0.81
Oman	0.90	68	0.73	0.91	0.88	0.68	0.82	0.90
Poland	0.93	75	0.80	0.90	0.92	0.74	0.90	0.91
Portugal	0.92	74	0.82	0.92	0.89	0.85	0.79	0.87
Qatar	0.96	84	0.87	0.95	0.93	0.85	0.92	0.95
Russian Federation	0.91	69	0.73	0.87	0.89	0.84	0.78	0.83
Saudi Arabia	0.93	74	0.83	0.93	0.92	0.66	0.86	0.92
Singapore	0.95	82	0.89	0.96	0.91	0.84	0.93	0.91
Slovak Republic	0.84	58	0.85	0.86	0.70	0.56	0.82	0.73
Slovenia	0.86	61	0.78	0.89	0.75	0.61	0.80	0.83
South Africa	0.93	75	0.76	0.88	0.90	0.86	0.88	0.91
Spain	0.90	70	0.82	0.90	0.88	0.74	0.79	0.88
Sweden	0.88	63	0.76	0.84	0.88	0.73	0.70	0.84
Trinidad and Tobago	0.92	72	0.81	0.90	0.87	0.74	0.87	0.88
United Arab Emirates	0.95	81	0.89	0.96	0.92	0.77	0.91	0.94
United States	0.97	87	0.91	0.96	0.94	0.89	0.95	0.95
Benchmarking Participants								
Buenos Aires, Argentina	0.90	66	0.81	0.90	0.83	0.62	0.83	0.84
Ontario, Canada	0.94	78	0.88	0.92	0.88	0.81	0.88	0.92
Quebec, Canada	0.89	65	0.78	0.91	0.82	0.58	0.86	0.86
Denmark (3)	0.88	63	0.82	0.88	0.82	0.69	0.77	0.78
Norway (4)	0.86	60	0.82	0.87	0.78	0.51	0.81	0.81
Moscow City, Russian Fed.	0.88	63	0.78	0.89	0.85	0.78	0.68	0.74
Eng/Afr/Zulu - RSA (5)	0.92	73	0.73	0.87	0.90	0.86	0.87	0.88
Andalusia, Spain	0.89	69	0.83	0.87	0.88	0.75	0.79	0.88
Madrid, Spain	0.81	58	0.73	0.82	0.85	0.77	0.66	0.73
Abu Dhabi, UAE	0.95	80	0.90	0.95	0.91	0.76	0.90	0.94
Dubai, UAE	0.95	81	0.90	0.96	0.92	0.74	0.93	0.95

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Schools Where Students Enter the Primary Grades with Early Literacy Skills* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.12	0.01	0.01
Austria	0.12	0.01	0.01
Azerbaijan	0.08	0.01	0.00
Bahrain	0.19	0.04	0.03
Belgium (Flemish)	0.05	0.00	0.00
Belgium (French)	0.12	0.01	0.02
Bulgaria	0.32	0.10	0.10
Canada	0.16	0.03	0.04
Chile	0.26	0.07	0.07
Chinese Taipei	0.05	0.00	0.02
Czech Republic	0.09	0.01	0.00
Denmark	0.09	0.01	0.01
Egypt	0.11	0.01	0.01
England	0.10	0.01	0.01
Finland	0.05	0.00	0.00
France	0.07	0.00	0.02
Georgia	0.07	0.00	0.00
Germany	0.17	0.03	0.01
Hong Kong SAR	0.09	0.01	0.02
Hungary	0.09	0.01	0.01
Iran, Islamic Rep. of	-0.06	0.00	0.01
Ireland	0.08	0.01	0.00
Israel	0.07	0.00	0.04
Italy	0.04	0.00	0.00
Kazakhstan	0.10	0.01	0.01
Kuwait	0.12	0.01	0.01
Latvia	0.08	0.01	0.01
Lithuania	0.15	0.02	0.01
Macao SAR	0.01	0.00	0.01
Malta	0.02	0.00	0.00
Morocco	0.26	0.07	0.06
Netherlands	0.07	0.00	0.01
New Zealand	0.17	0.03	0.05
Northern Ireland	0.12	0.01	0.02
Norway (5)	0.04	0.00	0.00
Oman	0.11	0.01	0.02
Poland	-0.02	0.00	0.00
Portugal	0.06	0.00	0.01
Qatar	0.20	0.04	0.04
Russian Federation	0.23	0.05	0.04
Saudi Arabia	0.01	0.00	0.01
Singapore	0.09	0.01	0.02
Slovak Republic	0.18	0.03	0.01
Slovenia	0.07	0.00	0.00
South Africa	-0.02	0.00	0.00
Spain	0.15	0.02	0.02
Sweden	0.18	0.03	0.02
Trinidad and Tobago	0.26	0.07	0.04
United Arab Emirates	0.34	0.12	0.11
United States	0.19	0.04	0.03
International Median	0.09	0.01	0.01
Benchmarking Participants			
Buenos Aires, Argentina	0.27	0.07	0.06
Ontario, Canada	0.15	0.02	0.01
Quebec, Canada	0.07	0.01	0.00
Denmark (3)	0.06	0.00	0.00
Norway (4)	0.09	0.01	0.01
Moscow City, Russian Fed.	0.18	0.03	0.03
Eng/Afr/Zulu - RSA (5)	0.07	0.00	0.03
Andalusia, Spain	0.14	0.02	0.02
Madrid, Spain	0.14	0.02	0.02
Abu Dhabi, UAE	0.26	0.07	0.06
Dubai, UAE	0.32	0.10	0.09

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Student Bullying Scale

The *Student Bullying* (SB) scale was created based on students' responses to how often they experienced the eight bullying behaviors described below.

Items in the PIRLS 2016 *Student Bullying* Scale

During this school year, how often have other students from your school done any of the following things to you (including through texting or the Internet)?

		Never	A few times a year	Once or twice a month	At least once a week
ASBG13A	1) Made fun of me or called me names	○	○	○	○
ASBG13B	2) Left me out of their games or activities	○	○	○	○
ASBG13C	3) Spread lies about me	○	○	○	○
ASBG13D	4) Stole something from me	○	○	○	○
ASBG13E	5) Hit or hurt me (*e.g., shoving, hitting, kicking*)	○	○	○	○
ASBG13F	6) Made me do things I didn't want to do	○	○	○	○
ASBG13G	7) Shared embarrassing information about me	○	○	○	○
ASBG13H	8) Threatened me	○	○	○	○

Almost Never | 9.5 | **About Monthly** | 7.9 | **About Weekly**

Item Parameters for the PIRLS 2016 *Student Bullying* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ASBG13A	0.47927	0.27533	-0.38571	0.11038	1.05
ASBG13B	0.28864	-0.00106	-0.07440	0.07546	1.16
ASBG13C	0.17327	0.02052	-0.09847	0.07795	0.96
ASBG13D	-0.31380	0.27095	-0.16042	-0.11053	1.12
ASBG13E	0.17865	0.00530	-0.18154	0.17624	1.02
ASBG13F	-0.30137	0.21016	0.06970	-0.27986	0.99
ASBG13G	-0.19155	0.16906	-0.08285	-0.08621	0.94
ASBG13H	-0.31311	0.37439	0.00776	-0.38215	0.94

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Student Bullying* Scale

Scale Transformation Constants	
A = 7.902851	Transformed Scale Score = 7.902851 + 1.812747 • Logit Scale Score
B = 1.812747	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Student Bullying* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.27794	
1	4.83400	
2	5.48907	
3	5.91604	
4	6.24033	
5	6.50472	
6	6.74041	
7	6.95212	
8	7.14715	
9	7.33106	
10	7.50803	
11	7.68141	
12	7.84988	7.9
13	8.02779	
14	8.20567	
15	8.39019	
16	8.58463	
17	8.79311	
18	9.02099	
19	9.27097	
20	9.56469	9.5
21	9.92468	
22	10.40216	
23	11.14752	
24	12.92197	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Student Bullying* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item							
			ASBG13A	ASBG13B	ASBG13C	ASBG13D	ASBG13E	ASBG13F	ASBG13G	ASBG13H
Australia	0.86	50	0.72	0.68	0.75	0.64	0.72	0.68	0.73	0.72
Austria	0.82	45	0.69	0.65	0.73	0.57	0.69	0.67	0.69	0.67
Azerbaijan	0.80	43	0.63	0.60	0.69	0.62	0.66	0.69	0.75	0.63
Bahrain	0.84	46	0.63	0.65	0.73	0.64	0.70	0.68	0.69	0.71
Belgium (Flemish)	0.84	47	0.72	0.68	0.74	0.53	0.72	0.67	0.70	0.67
Belgium (French)	0.83	45	0.69	0.61	0.72	0.60	0.70	0.66	0.70	0.72
Bulgaria	0.81	44	0.66	0.53	0.74	0.52	0.73	0.67	0.73	0.70
Canada	0.85	49	0.71	0.66	0.75	0.62	0.72	0.68	0.72	0.72
Chile	0.84	49	0.69	0.69	0.72	0.64	0.74	0.74	0.66	0.73
Chinese Taipei	0.84	47	0.72	0.69	0.73	0.51	0.69	0.67	0.76	0.71
Czech Republic	0.82	44	0.72	0.55	0.73	0.58	0.66	0.66	0.75	0.65
Denmark	0.84	48	0.70	0.68	0.74	0.54	0.72	0.71	0.71	0.72
Egypt	0.84	49	0.66	0.65	0.73	0.66	0.72	0.74	0.74	0.67
England	0.85	49	0.73	0.68	0.77	0.56	0.72	0.68	0.71	0.73
Finland	0.84	48	0.73	0.67	0.75	0.54	0.72	0.67	0.74	0.71
France	0.80	43	0.68	0.61	0.72	0.53	0.67	0.62	0.68	0.69
Georgia	0.82	47	0.69	0.57	0.73	0.66	0.69	0.72	0.72	0.72
Germany	0.82	45	0.65	0.64	0.72	0.56	0.71	0.67	0.71	0.67
Hong Kong SAR	0.83	46	0.65	0.65	0.76	0.54	0.66	0.74	0.63	0.74
Hungary	0.79	41	0.64	0.64	0.69	0.52	0.65	0.64	0.67	0.67
Iran, Islamic Rep. of	0.82	45	0.66	0.57	0.72	0.62	0.71	0.66	0.71	0.72
Ireland	0.84	47	0.71	0.67	0.75	0.58	0.69	0.66	0.70	0.70
Israel	-	-	-	-	-	-	-	-	-	-
Italy	0.81	42	0.69	0.63	0.70	0.54	0.64	0.63	0.70	0.65
Kazakhstan	0.81	45	0.65	0.58	0.72	0.62	0.69	0.65	0.74	0.68
Kuwait	0.77	39	0.61	0.62	0.64	0.56	0.64	0.61	0.67	0.63
Latvia	0.84	47	0.73	0.61	0.76	0.54	0.74	0.66	0.75	0.69
Lithuania	0.81	44	0.69	0.59	0.72	0.57	0.68	0.64	0.71	0.68
Macao SAR	0.78	40	0.57	0.59	0.71	0.55	0.63	0.68	0.58	0.71
Malta	0.86	50	0.72	0.66	0.76	0.64	0.71	0.72	0.74	0.73
Morocco	0.80	41	0.61	0.62	0.69	0.60	0.66	0.63	0.68	0.66
Netherlands	0.83	46	0.73	0.69	0.73	0.54	0.70	0.67	0.68	0.68
New Zealand	0.87	52	0.73	0.70	0.76	0.66	0.73	0.71	0.76	0.71
Northern Ireland	0.83	47	0.73	0.68	0.74	0.57	0.69	0.66	0.70	0.69
Norway (5)	0.84	48	0.74	0.69	0.76	0.52	0.70	0.68	0.72	0.71
Oman	0.82	44	0.64	0.57	0.70	0.65	0.68	0.67	0.71	0.67
Poland	0.85	50	0.70	0.69	0.76	0.58	0.72	0.69	0.79	0.73
Portugal	0.83	46	0.69	0.62	0.73	0.56	0.74	0.65	0.70	0.71
Qatar	0.85	49	0.67	0.61	0.73	0.67	0.72	0.73	0.74	0.73
Russian Federation	0.79	42	0.69	0.44	0.72	0.50	0.72	0.61	0.76	0.66
Saudi Arabia	0.87	53	0.68	0.66	0.74	0.67	0.75	0.75	0.78	0.75
Singapore	0.83	47	0.65	0.65	0.73	0.59	0.70	0.68	0.72	0.72
Slovak Republic	0.82	46	0.70	0.59	0.74	0.55	0.68	0.68	0.73	0.71
Slovenia	0.84	47	0.71	0.63	0.77	0.57	0.69	0.66	0.76	0.69
South Africa	0.79	40	0.60	0.58	0.65	0.56	0.65	0.68	0.68	0.64
Spain	0.82	44	0.66	0.63	0.71	0.55	0.69	0.67	0.65	0.71
Sweden	0.84	48	0.71	0.67	0.77	0.59	0.71	0.68	0.70	0.71
Trinidad and Tobago	0.80	42	0.63	0.56	0.69	0.60	0.70	0.59	0.72	0.66
United Arab Emirates	0.84	47	0.65	0.62	0.72	0.66	0.72	0.71	0.71	0.72
United States	0.85	49	0.72	0.65	0.76	0.66	0.70	0.65	0.73	0.72
Benchmarking Participants										
Buenos Aires, Argentina	0.81	47	0.66	0.66	0.74	0.61	0.70	-	0.72	0.70
Ontario, Canada	0.85	49	0.70	0.67	0.76	0.63	0.72	0.69	0.71	0.71
Quebec, Canada	0.84	47	0.74	0.63	0.74	0.55	0.70	0.64	0.74	0.73
Denmark (3)	0.83	47	0.69	0.63	0.73	0.52	0.71	0.69	0.70	0.76
Norway (4)	0.85	48	0.72	0.67	0.76	0.56	0.71	0.70	0.71	0.73
Moscow City, Russian Fed.	0.83	46	0.73	0.51	0.73	0.57	0.73	0.66	0.77	0.70
Eng/Afr/Zulu - RSA (5)	0.78	39	0.59	0.55	0.65	0.52	0.67	0.66	0.69	0.67
Andalusia, Spain	0.80	42	0.64	0.62	0.70	0.54	0.71	0.65	0.64	0.69
Madrid, Spain	0.80	42	0.65	0.65	0.69	0.51	0.67	0.65	0.64	0.70
Abu Dhabi, UAE	0.85	48	0.64	0.62	0.72	0.66	0.71	0.74	0.72	0.73
Dubai, UAE	0.85	48	0.66	0.65	0.75	0.64	0.71	0.69	0.74	0.70

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Student Bullying* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.13	0.02	0.03
Austria	0.13	0.02	0.03
Azerbaijan	0.13	0.02	0.03
Bahrain	0.22	0.05	0.05
Belgium (Flemish)	0.10	0.01	0.02
Belgium (French)	0.09	0.01	0.01
Bulgaria	0.12	0.01	0.02
Canada	0.13	0.02	0.03
Chile	0.19	0.04	0.06
Chinese Taipei	0.11	0.01	0.02
Czech Republic	0.11	0.01	0.02
Denmark	0.06	0.00	0.01
Egypt	0.10	0.01	0.01
England	0.14	0.02	0.03
Finland	0.12	0.01	0.02
France	0.14	0.02	0.03
Georgia	0.18	0.03	0.04
Germany	0.15	0.02	0.04
Hong Kong SAR	0.11	0.01	0.02
Hungary	0.12	0.02	0.03
Iran, Islamic Rep. of	0.01	0.00	0.00
Ireland	0.18	0.03	0.04
Israel	-	-	-
Italy	0.07	0.00	0.01
Kazakhstan	0.10	0.01	0.01
Kuwait	0.12	0.01	0.02
Latvia	0.18	0.03	0.04
Lithuania	0.19	0.04	0.04
Macao SAR	0.11	0.01	0.02
Malta	0.16	0.02	0.04
Morocco	0.15	0.02	0.03
Netherlands	0.07	0.00	0.01
New Zealand	0.17	0.03	0.04
Northern Ireland	0.13	0.02	0.03
Norway (5)	0.09	0.01	0.02
Oman	0.15	0.02	0.03
Poland	0.18	0.03	0.04
Portugal	0.10	0.01	0.01
Qatar	0.20	0.04	0.06
Russian Federation	0.11	0.01	0.02
Saudi Arabia	0.25	0.06	0.08
Singapore	0.18	0.03	0.04
Slovak Republic	0.14	0.02	0.03
Slovenia	0.09	0.01	0.02
South Africa	0.17	0.03	0.04
Spain	0.14	0.02	0.03
Sweden	0.11	0.01	0.02
Trinidad and Tobago	0.12	0.01	0.02
United Arab Emirates	0.22	0.05	0.06
United States	0.13	0.02	0.03
International Median	0.13	0.02	0.03
Benchmarking Participants			
Buenos Aires, Argentina	0.14	0.02	0.03
Ontario, Canada	0.14	0.02	0.03
Quebec, Canada	0.08	0.01	0.01
Denmark (3)	0.12	0.01	0.02
Norway (4)	0.07	0.00	0.01
Moscow City, Russian Fed.	0.15	0.02	0.02
Eng/Afr/Zulu - RSA (5)	0.19	0.04	0.05
Andalusia, Spain	0.11	0.01	0.02
Madrid, Spain	0.09	0.01	0.02
Abu Dhabi, UAE	0.23	0.05	0.07
Dubai, UAE	0.16	0.03	0.05

A dash (–) indicates comparable data not available.

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Students Confident in Reading Scale

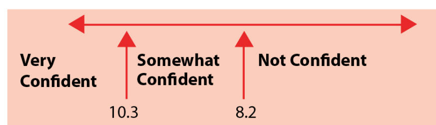
The *Students Confident in Reading* (SCR) scale was created based on students' degree of agreement with the six statements described below.

Items in the PIRLS 2016 *Students Confident in Reading* Scale

How well do you read? Tell how much you agree with each of these statements.

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBR07A	1) I usually do well in reading	◯	◯	◯	◯
ASBR07B	2) Reading is easy for me	◯	◯	◯	◯
ASBR07C*	3) I have trouble reading stories with difficult words*	◯	◯	◯	◯
ASBR07D*	4) Reading is harder for me than for many of my classmates*	◯	◯	◯	◯
ASBR07E*	5) Reading is harder for me than any other subject*	◯	◯	◯	◯
ASBR07F*	6) I am just not good at reading*	◯	◯	◯	◯

* Reverse coded



Item Parameters for the PIRLS 2016 *Students Confident in Reading* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ASBR07A	-0.52312	-0.32359	-0.68416	1.00775	1.11
ASBR07B	-0.60120	-0.48024	-0.42689	0.90713	1.03
* ASBR07C	1.01425	-0.93819	0.50887	0.42932	1.22
* ASBR07D	0.25813	-0.38905	0.22763	0.16142	0.94
* ASBR07E	-0.00713	-0.11719	0.22036	-0.10317	0.92
* ASBR07F	-0.14093	-0.05955	0.23700	-0.17745	0.90

* Reverse coded

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Students Confident in Reading* Scale

Scale Transformation Constants	
A = 8.137507	Transformed Scale Score = 8.137507 + 1.753646 • Logit Scale Score
B = 1.753646	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Students Confident in Reading* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	2.95794	
1	4.60752	
2	5.40368	
3	5.97279	
4	6.43307	
5	6.82953	
6	7.18533	
7	7.50982	
8	7.81385	
9	8.10473	8.2
10	8.40103	
11	8.70504	
12	9.03187	
13	9.39594	
14	9.81182	
15	10.31185	10.3
16	10.93952	
17	11.79542	
18	13.47027	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Students Confident in Reading* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item					
			ASBR07A	ASBR07B	ASBR07C*	ASBR07D*	ASBR07E*	ASBR07F*
Australia	0.81	52	0.67	0.73	0.62	0.79	0.75	0.77
Austria	0.81	53	0.71	0.74	0.59	0.76	0.76	0.79
Azerbaijan	0.68	40	0.11	0.25	0.51	0.84	0.84	0.81
Bahrain	0.69	40	0.19	0.40	0.62	0.76	0.81	0.75
Belgium (Flemish)	0.83	56	0.75	0.80	0.57	0.76	0.75	0.83
Belgium (French)	0.75	45	0.59	0.60	0.55	0.74	0.75	0.77
Bulgaria	0.80	53	0.71	0.73	0.49	0.78	0.80	0.80
Canada	0.81	53	0.69	0.75	0.57	0.78	0.78	0.78
Chile	0.69	39	0.38	0.48	0.57	0.71	0.75	0.78
Chinese Taipei	0.78	49	0.60	0.66	0.45	0.81	0.81	0.80
Czech Republic	0.79	50	0.65	0.73	0.50	0.78	0.75	0.80
Denmark	0.83	56	0.77	0.80	0.58	0.75	0.78	0.79
Egypt	0.72	42	0.32	0.60	0.53	0.77	0.80	0.76
England	0.82	54	0.68	0.74	0.61	0.80	0.78	0.79
Finland	0.80	53	0.72	0.75	0.57	0.79	0.76	0.78
France	0.78	48	0.68	0.62	0.57	0.77	0.72	0.79
Georgia	0.75	45	0.45	0.52	0.49	0.81	0.82	0.82
Germany	0.82	54	0.71	0.74	0.58	0.78	0.78	0.80
Hong Kong SAR	0.80	50	0.62	0.67	0.57	0.79	0.80	0.76
Hungary	0.82	55	0.76	0.75	0.52	0.80	0.77	0.82
Iran, Islamic Rep. of	0.70	41	0.40	0.46	0.59	0.74	0.79	0.75
Ireland	0.82	54	0.72	0.78	0.61	0.79	0.75	0.76
Israel	0.75	46	0.42	0.63	0.49	0.80	0.80	0.82
Italy	0.75	45	0.63	0.63	0.61	0.75	0.67	0.73
Kazakhstan	0.77	48	0.37	0.44	0.67	0.85	0.85	0.79
Kuwait	0.65	37	0.44	0.57	0.42	0.69	0.73	0.73
Latvia	0.80	52	0.73	0.76	0.51	0.79	0.76	0.74
Lithuania	0.80	52	0.69	0.74	0.54	0.80	0.78	0.75
Macao SAR	0.75	45	0.61	0.63	0.46	0.76	0.77	0.75
Malta	0.71	42	0.55	0.53	0.49	0.79	0.73	0.74
Morocco	0.56	34	-0.08	0.32	0.58	0.74	0.76	0.71
Netherlands	0.85	59	0.73	0.82	0.58	0.79	0.80	0.84
New Zealand	0.76	46	0.55	0.63	0.55	0.79	0.78	0.74
Northern Ireland	0.83	55	0.70	0.74	0.65	0.78	0.77	0.79
Norway (5)	0.82	54	0.71	0.77	0.64	0.76	0.80	0.74
Oman	0.67	38	0.33	0.40	0.55	0.75	0.78	0.72
Poland	0.83	56	0.68	0.73	0.65	0.80	0.80	0.79
Portugal	0.72	44	0.71	0.65	0.56	0.66	0.57	0.79
Qatar	0.72	42	0.19	0.36	0.65	0.80	0.83	0.80
Russian Federation	0.79	51	0.71	0.72	0.54	0.80	0.74	0.73
Saudi Arabia	0.53	41	-0.41	0.26	0.64	0.73	0.83	0.75
Singapore	0.79	49	0.65	0.70	0.57	0.77	0.75	0.76
Slovak Republic	0.81	53	0.68	0.74	0.54	0.79	0.76	0.82
Slovenia	0.83	55	0.73	0.77	0.66	0.78	0.77	0.74
South Africa	0.58	35	-0.13	-0.07	0.61	0.74	0.77	0.74
Spain	0.68	40	0.61	0.60	0.41	0.70	0.67	0.73
Sweden	0.82	55	0.69	0.75	0.65	0.77	0.77	0.79
Trinidad and Tobago	0.76	47	0.63	0.66	0.49	0.75	0.77	0.77
United Arab Emirates	0.72	42	0.39	0.47	0.61	0.76	0.79	0.75
United States	0.79	49	0.62	0.67	0.57	0.80	0.76	0.76
Benchmarking Participants								
Buenos Aires, Argentina	0.65	37	0.47	0.43	0.46	0.77	0.65	0.78
Ontario, Canada	0.82	54	0.70	0.76	0.57	0.79	0.79	0.79
Quebec, Canada	0.82	54	0.70	0.75	0.58	0.77	0.78	0.80
Denmark (3)	0.80	51	0.71	0.75	0.60	0.74	0.73	0.77
Norway (4)	0.79	50	0.65	0.73	0.61	0.74	0.77	0.74
Moscow City, Russian Fed.	0.80	52	0.74	0.77	0.53	0.80	0.76	0.70
Eng/Afr/Zulu - RSA (5)	0.68	39	0.23	0.30	0.56	0.80	0.80	0.77
Andalusia, Spain	0.66	38	0.59	0.51	0.42	0.71	0.68	0.73
Madrid, Spain	0.67	39	0.63	0.57	0.42	0.71	0.65	0.72
Abu Dhabi, UAE	0.71	41	0.38	0.46	0.61	0.75	0.78	0.75
Dubai, UAE	0.75	45	0.51	0.55	0.58	0.77	0.78	0.76

*Reverse coded

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Students Confident in Reading* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.52	0.27	0.28
Austria	0.41	0.17	0.16
Azerbaijan	0.31	0.10	0.11
Bahrain	0.45	0.20	0.22
Belgium (Flemish)	0.33	0.11	0.10
Belgium (French)	0.43	0.18	0.19
Bulgaria	0.42	0.17	0.18
Canada	0.45	0.20	0.21
Chile	0.39	0.15	0.17
Chinese Taipei	0.40	0.16	0.17
Czech Republic	0.40	0.16	0.17
Denmark	0.50	0.25	0.23
Egypt	0.39	0.15	0.16
England	0.49	0.24	0.23
Finland	0.42	0.18	0.17
France	0.41	0.17	0.19
Georgia	0.38	0.15	0.16
Germany	0.40	0.16	0.17
Hong Kong SAR	0.39	0.15	0.14
Hungary	0.49	0.24	0.22
Iran, Islamic Rep. of	0.43	0.18	0.19
Ireland	0.44	0.19	0.19
Israel	0.46	0.21	0.24
Italy	0.30	0.09	0.11
Kazakhstan	0.23	0.05	0.07
Kuwait	0.36	0.13	0.14
Latvia	0.41	0.17	0.17
Lithuania	0.47	0.22	0.20
Macao SAR	0.37	0.14	0.13
Malta	0.46	0.21	0.21
Morocco	0.42	0.17	0.19
Netherlands	0.37	0.14	0.13
New Zealand	0.50	0.25	0.26
Northern Ireland	0.46	0.21	0.23
Norway (5)	0.46	0.21	0.19
Oman	0.43	0.18	0.19
Poland	0.40	0.16	0.20
Portugal	0.39	0.16	0.17
Qatar	0.43	0.18	0.21
Russian Federation	0.41	0.17	0.19
Saudi Arabia	0.33	0.11	0.12
Singapore	0.49	0.24	0.24
Slovak Republic	0.42	0.18	0.19
Slovenia	0.43	0.19	0.20
South Africa	0.39	0.15	0.15
Spain	0.39	0.15	0.16
Sweden	0.41	0.17	0.17
Trinidad and Tobago	0.53	0.29	0.30
United Arab Emirates	0.48	0.23	0.25
United States	0.44	0.20	0.19
International Median	0.42	0.17	0.19
Benchmarking Participants			
Buenos Aires, Argentina	0.43	0.19	0.19
Ontario, Canada	0.45	0.20	0.21
Quebec, Canada	0.43	0.19	0.19
Denmark (3)	0.47	0.22	0.21
Norway (4)	0.45	0.20	0.20
Moscow City, Russian Fed.	0.40	0.16	0.16
Eng/Afr/Zulu - RSA (5)	0.43	0.19	0.18
Andalusia, Spain	0.42	0.18	0.20
Madrid, Spain	0.37	0.14	0.14
Abu Dhabi, UAE	0.50	0.25	0.27
Dubai, UAE	0.44	0.20	0.22

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Students Engaged in Reading Lessons Scale

The *Students Engaged in Reading Lessons* (ERL) scale was created based on students' degree of agreement with the nine statements described below.

Items in the PIRLS 2016 *Students Engaged in Reading Lessons* Scale

How much do you agree with these statements about your reading lessons?

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBR01A	1) I like what I read about in school	○	○	○	○
ASBR01B	2) My teacher gives me interesting things to read	○	○	○	○
ASBR01C	3) I know what my teacher expects me to do	○	○	○	○
ASBR01D	4) My teacher is easy to understand	○	○	○	○
ASBR01E	5) I am interested in what my teacher says	○	○	○	○
ASBR01F	6) My teacher encourages me to say what I think about what I have read	○	○	○	○
ASBR01G	7) My teacher lets me show what I have learned	○	○	○	○
ASBR01H	8) My teacher does a variety of things to help us learn	○	○	○	○
ASBR01I	9) My teacher tells me how to do better when I make a mistake	○	○	○	○

Very Engaged | 9.5 | Somewhat Engaged | 7.1 | Less than Engaged

Item Parameters for the PIRLS 2016 *Students Engaged in Reading Lessons* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ASBR01A	0.21864	-0.59198	-0.74411	1.33609	1.06
ASBR01B	0.25459	-0.68288	-0.46758	1.15046	0.94
ASBR01C	0.06439	-0.36656	-0.61903	0.98559	1.19
ASBR01D	-0.12990	-0.58489	-0.54402	1.12891	1.04
ASBR01E	-0.02167	-0.56053	-0.56022	1.12075	0.95
ASBR01F	0.35783	-0.55558	-0.43518	0.99076	1.04
ASBR01G	0.13972	-0.56265	-0.42458	0.98723	1.01
ASBR01H	-0.53069	-0.08820	-0.47557	0.56377	0.98
ASBR01I	-0.35291	-0.16188	-0.43241	0.59429	1.00

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Students Engaged in Reading Lessons* Scale

Scale Transformation Constants	
A = 7.347685	Transformed Scale Score = 7.347685 + 1.442440 • Logit Scale Score
B = 1.442440	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Students Engaged in Reading Lessons* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	2.54080	
1	3.94030	
2	4.57040	
3	4.98984	
4	5.31080	
5	5.57732	
6	5.80694	
7	6.01513	
8	6.20800	
9	6.39051	
10	6.56630	
11	6.73824	
12	6.91081	
13	7.08114	7.1
14	7.25473	
15	7.43394	
16	7.62129	
17	7.81965	
18	8.03231	
19	8.26228	
20	8.51711	
21	8.80348	
22	9.13146	
23	9.51587	9.5
24	9.98157	
25	10.57734	
26	11.42874	
27	13.13080	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Students Engaged in Reading Lessons* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item								
			ASBR01A	ASBR01B	ASBR01C	ASBR01D	ASBR01E	ASBR01F	ASBR01G	ASBR01H	ASBR01I
Australia	0.84	44	0.61	0.72	0.59	0.66	0.75	0.67	0.65	0.67	0.65
Austria	0.85	46	0.63	0.73	0.61	0.66	0.72	0.72	0.70	0.70	0.64
Azerbaijan	0.83	44	0.55	0.70	0.55	0.63	0.73	0.69	0.69	0.68	0.71
Bahrain	0.84	44	0.59	0.67	0.61	0.66	0.72	0.69	0.69	0.67	0.64
Belgium (Flemish)	0.80	39	0.58	0.69	0.51	0.60	0.67	0.63	0.68	0.63	0.63
Belgium (French)	0.84	45	0.63	0.74	0.68	0.67	0.71	0.66	0.57	0.70	0.64
Bulgaria	0.85	46	0.64	0.76	0.55	0.62	0.63	0.70	0.74	0.71	0.72
Canada	0.84	44	0.60	0.71	0.61	0.67	0.73	0.68	0.65	0.65	0.65
Chile	0.88	50	0.63	0.74	0.69	0.70	0.74	0.72	0.72	0.72	0.72
Chinese Taipei	0.89	52	0.61	0.74	0.73	0.68	0.75	0.75	0.76	0.76	0.74
Czech Republic	0.85	45	0.65	0.70	0.57	0.67	0.69	0.69	0.65	0.69	0.70
Denmark	0.85	46	0.67	0.74	0.58	0.59	0.71	0.74	0.72	0.72	0.63
Egypt	0.79	40	0.63	0.60	0.46	0.66	0.71	0.71	0.70	0.58	0.60
England	0.84	44	0.63	0.71	0.62	0.65	0.72	0.65	0.65	0.67	0.66
Finland	0.86	48	0.62	0.72	0.61	0.72	0.73	0.70	0.70	0.72	0.70
France	0.83	42	0.67	0.73	0.55	0.58	0.72	0.65	0.67	0.62	0.64
Georgia	0.72	35	0.56	0.69	0.49	0.64	0.47	0.47	0.69	0.66	0.64
Germany	0.85	46	0.63	0.75	0.61	0.73	0.69	0.71	0.67	0.66	0.66
Hong Kong SAR	0.89	52	0.66	0.78	0.70	0.68	0.78	0.74	0.73	0.72	0.71
Hungary	0.86	48	0.65	0.71	0.64	0.69	0.74	0.70	0.66	0.72	0.70
Iran, Islamic Rep. of	0.82	42	0.64	0.65	0.52	0.68	0.73	0.63	0.70	0.64	0.64
Ireland	0.83	44	0.64	0.76	0.59	0.64	0.75	0.64	0.62	0.65	0.63
Israel	0.87	49	0.67	0.77	0.59	0.66	0.77	0.71	0.71	0.69	0.71
Italy	0.78	37	0.66	0.71	0.49	0.57	0.69	0.59	0.59	0.58	0.60
Kazakhstan	0.80	40	0.54	0.66	0.59	0.64	0.66	0.68	0.67	0.57	0.62
Kuwait	0.79	38	0.59	0.67	0.52	0.59	0.67	0.63	0.65	0.60	0.60
Latvia	0.83	43	0.62	0.71	0.60	0.67	0.71	0.62	0.66	0.66	0.66
Lithuania	0.81	40	0.62	0.71	0.60	0.62	0.64	0.65	0.64	0.60	0.62
Macao SAR	0.85	46	0.68	0.74	0.65	0.66	0.71	0.69	0.69	0.66	0.62
Malta	0.81	41	0.61	0.69	0.64	0.55	0.69	0.65	0.65	0.62	0.65
Morocco	0.76	35	0.53	0.56	0.54	0.61	0.64	0.64	0.67	0.56	0.55
Netherlands	0.81	41	0.60	0.69	0.54	0.63	0.71	0.61	0.67	0.65	0.62
New Zealand	0.83	43	0.63	0.73	0.60	0.65	0.73	0.65	0.64	0.67	0.63
Northern Ireland	0.83	42	0.61	0.72	0.57	0.66	0.71	0.65	0.66	0.64	0.64
Norway (5)	0.83	43	0.62	0.71	0.57	0.66	0.69	0.72	0.66	0.63	0.62
Oman	0.83	44	0.59	0.68	0.51	0.67	0.72	0.69	0.69	0.70	0.66
Poland	0.86	48	0.66	0.75	0.66	0.63	0.74	0.68	0.71	0.69	0.68
Portugal	0.79	39	0.63	0.70	0.49	0.60	0.67	0.63	0.64	0.58	0.64
Qatar	0.87	49	0.65	0.73	0.68	0.71	0.74	0.69	0.69	0.72	0.68
Russian Federation	0.81	40	0.56	0.68	0.59	0.65	0.68	0.62	0.64	0.65	0.65
Saudi Arabia	0.82	42	0.48	0.64	0.56	0.67	0.70	0.71	0.71	0.69	0.66
Singapore	0.85	45	0.63	0.72	0.61	0.67	0.74	0.68	0.65	0.68	0.68
Slovak Republic	0.84	44	0.64	0.73	0.59	0.65	0.71	0.66	0.66	0.69	0.64
Slovenia	0.85	46	0.65	0.73	0.64	0.70	0.60	0.68	0.68	0.72	0.70
South Africa	0.84	44	0.62	0.70	0.65	0.64	0.62	0.66	0.70	0.69	0.66
Spain	0.80	40	0.61	0.70	0.47	0.57	0.70	0.66	0.65	0.66	0.63
Sweden	0.86	47	0.63	0.72	0.61	0.70	0.71	0.72	0.69	0.71	0.65
Trinidad and Tobago	0.83	42	0.60	0.69	0.61	0.66	0.73	0.66	0.62	0.63	0.64
United Arab Emirates	0.85	45	0.62	0.72	0.61	0.68	0.71	0.68	0.69	0.69	0.64
United States	0.86	47	0.60	0.74	0.62	0.69	0.76	0.71	0.67	0.67	0.68
Benchmarking Participants											
Buenos Aires, Argentina	0.83	44	0.60	0.68	0.55	0.56	0.70	0.67	0.72	0.72	0.71
Ontario, Canada	0.84	45	0.61	0.71	0.63	0.70	0.74	0.70	0.64	0.65	0.65
Quebec, Canada	0.83	43	0.54	0.71	0.60	0.65	0.75	0.64	0.69	0.65	0.65
Denmark (3)	0.83	43	0.66	0.73	0.59	0.51	0.73	0.70	0.72	0.65	0.59
Norway (4)	0.82	42	0.63	0.72	0.54	0.62	0.71	0.70	0.63	0.59	0.62
Moscow City, Russian Fed.	0.83	43	0.61	0.73	0.59	0.66	0.73	0.60	0.63	0.67	0.66
Eng/Afr/Zulu - RSA (5)	0.81	40	0.63	0.68	0.63	0.61	0.62	0.66	0.64	0.62	0.63
Andalusia, Spain	0.79	39	0.58	0.67	0.46	0.55	0.68	0.63	0.68	0.67	0.64
Madrid, Spain	0.81	41	0.61	0.70	0.51	0.61	0.65	0.67	0.67	0.66	0.66
Abu Dhabi, UAE	0.85	45	0.61	0.71	0.59	0.69	0.71	0.70	0.68	0.72	0.64
Dubai, UAE	0.84	44	0.63	0.71	0.62	0.63	0.71	0.66	0.69	0.66	0.64

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Students Engaged in Reading Lessons* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement (r)	Pearson's Correlation with Reading Achievement (r^2)	Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
Australia	0.02	0.00	0.00
Austria	0.02	0.00	0.00
Azerbaijan	0.14	0.02	0.02
Bahrain	0.17	0.03	0.03
Belgium (Flemish)	0.04	0.00	0.00
Belgium (French)	-0.01	0.00	0.00
Bulgaria	0.06	0.00	0.01
Canada	0.09	0.01	0.01
Chile	0.12	0.01	0.03
Chinese Taipei	0.09	0.01	0.01
Czech Republic	-0.06	0.00	0.01
Denmark	0.15	0.02	0.03
Egypt	0.15	0.02	0.02
England	0.04	0.00	0.01
Finland	0.08	0.01	0.01
France	-0.01	0.00	0.00
Georgia	0.14	0.02	0.03
Germany	0.12	0.01	0.03
Hong Kong SAR	0.11	0.01	0.02
Hungary	0.05	0.00	0.00
Iran, Islamic Rep. of	0.08	0.01	0.02
Ireland	0.03	0.00	0.00
Israel	-0.09	0.01	0.00
Italy	0.06	0.00	0.01
Kazakhstan	0.04	0.00	0.00
Kuwait	0.13	0.02	0.03
Latvia	0.03	0.00	0.00
Lithuania	0.09	0.01	0.02
Macao SAR	0.10	0.01	0.01
Malta	0.19	0.04	0.05
Morocco	0.10	0.01	0.01
Netherlands	0.05	0.00	0.01
New Zealand	-0.01	0.00	0.00
Northern Ireland	-0.01	0.00	0.00
Norway (5)	0.05	0.00	0.00
Oman	0.15	0.02	0.04
Poland	-0.04	0.00	0.01
Portugal	0.09	0.01	0.01
Qatar	0.16	0.03	0.04
Russian Federation	-0.01	0.00	0.00
Saudi Arabia	0.16	0.03	0.04
Singapore	0.03	0.00	0.01
Slovak Republic	-0.08	0.01	0.00
Slovenia	-0.03	0.00	0.00
South Africa	0.16	0.03	0.04
Spain	0.03	0.00	0.00
Sweden	0.04	0.00	0.01
Trinidad and Tobago	0.09	0.01	0.01
United Arab Emirates	0.17	0.03	0.04
United States	0.08	0.01	0.01
International Median	0.07	0.01	0.01
Benchmarking Participants			
Buenos Aires, Argentina	0.02	0.00	0.00
Ontario, Canada	0.11	0.01	0.02
Quebec, Canada	0.06	0.00	0.01
Denmark (3)	0.05	0.00	0.01
Norway (4)	0.04	0.00	0.01
Moscow City, Russian Fed.	0.02	0.00	0.00
Eng/Afr/Zulu - RSA (5)	0.04	0.00	0.01
Andalusia, Spain	0.04	0.00	0.00
Madrid, Spain	-0.02	0.00	0.00
Abu Dhabi, UAE	0.16	0.03	0.04
Dubai, UAE	0.13	0.02	0.04

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Students Like Reading Scale

The *Students Like Reading* (SLR) scale was created based on students' responses to the ten items described below.

Items in the PIRLS 2016 *Students Like Reading* Scale

What do you think about reading? Tell how much you agree with each of these statements.

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBR06A	1) I like talking about what I read with other people	◯	◯	◯	◯
ASBR06B	2) I would be happy if someone gave me a book as a present	◯	◯	◯	◯
ASBR06C*	3) I think reading is boring*	◯	◯	◯	◯
ASBR06D	4) I would like to have more time for reading	◯	◯	◯	◯
ASBR06E	5) I enjoy reading	◯	◯	◯	◯
ASBR06F	6) I learn a lot from reading	◯	◯	◯	◯
ASBR06G	7) I like to read things that make me think	◯	◯	◯	◯
ASBR06H	8) I like it when a book helps me imagine other worlds	◯	◯	◯	◯

* Reverse coded

Very Much Like Reading | 10.3 | **Somewhat Like Reading** | 8.3 | **Do Not Like Reading**

How often do you do these things outside of school?

		Every day or almost every day	Once or twice a week	Once or twice a month	Never or almost never
ASBR05A	1) I read for fun	◯	◯	◯	◯
ASBR05B	2) I read to find out about things I want to learn	◯	◯	◯	◯

Very Much Like Reading | 10.3 | **Somewhat Like Reading** | 8.3 | **Do Not Like Reading**

Item Parameters for the PIRLS 2016 *Students Like Reading* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ASBR06A	0.34442	-0.46997	-0.35824	0.82821	1.12
ASBR06B	-0.11815	-0.20564	-0.36880	0.57444	0.92
* ASBR06C	0.07123	-0.25296	0.13935	0.11361	1.27
ASBR06D	0.31398	-0.56008	-0.11801	0.67809	0.89
ASBR06E	-0.17844	-0.19317	-0.35192	0.54509	0.72
ASBR06F	-0.57533	-0.50890	-0.23851	0.74741	0.90
ASBR06G	-0.11502	-0.39988	-0.24307	0.64295	0.95
ASBR06H	-0.45737	0.03019	-0.28447	0.25428	1.02
ASBR05A	0.40723	-0.22541	-0.47027	0.69568	1.18
ASBR05B	0.30745	-0.62547	-0.19966	0.82513	1.13

* Reverse coded

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Students Like Reading* Scale

Scale Transformation Constants	
A = 8.281596	Transformed Scale Score = 8.281596 + 1.704604 • Logit Scale Score
B = 1.704604	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Students Like Reading* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	2.55082	
1	4.26783	
2	5.03823	
3	5.54349	
4	5.92440	
5	6.23597	
6	6.50253	
7	6.73682	
8	6.95316	
9	7.15371	
10	7.34241	
11	7.52243	
12	7.69637	
13	7.86639	
14	8.03431	
15	8.20182	8.3
16	8.37054	
17	8.54211	
18	8.71838	
19	8.90141	
20	9.09362	
21	9.29790	
22	9.51573	
23	9.75594	
24	10.02472	
25	10.33319	10.3
26	10.69833	
27	11.15205	
28	11.75519	
29	12.66304	
30	14.58422	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Students Like Reading* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item									
			ASBR06A	ASBR06B	ASBR06C*	ASBR06D	ASBR06E	ASBR06F	ASBR06G	ASBR06H	ASBR05A	ASBR05B
Australia	0.88	49	0.53	0.73	0.77	0.78	0.84	0.68	0.71	0.66	0.67	0.55
Austria	0.86	45	0.54	0.67	0.65	0.71	0.83	0.66	0.62	0.67	0.74	0.58
Azerbaijan	0.68	33	0.34	0.63	0.24	0.69	0.75	0.74	0.69	0.67	0.28	0.42
Bahrain	0.83	41	0.55	0.66	0.40	0.75	0.80	0.73	0.70	0.64	0.48	0.56
Belgium (Flemish)	0.87	47	0.56	0.73	0.76	0.77	0.85	0.66	0.69	0.60	0.69	0.47
Belgium (French)	0.86	45	0.56	0.73	0.53	0.74	0.81	0.68	0.71	0.63	0.63	0.61
Bulgaria	0.89	50	0.66	0.72	0.55	0.78	0.83	0.72	0.73	0.68	0.71	0.66
Canada	0.86	45	0.52	0.70	0.70	0.77	0.83	0.66	0.68	0.61	0.63	0.55
Chile	0.88	49	0.64	0.75	0.47	0.76	0.85	0.73	0.80	0.69	0.60	0.67
Chinese Taipei	0.89	51	0.59	0.62	0.64	0.84	0.86	0.79	0.72	0.67	0.66	0.67
Czech Republic	0.88	47	0.63	0.72	0.69	0.70	0.83	0.69	0.62	0.63	0.71	0.64
Denmark	0.85	44	0.62	0.71	0.69	0.71	0.84	0.64	0.75	0.66	0.50	0.48
Egypt	0.84	44	0.64	0.67	0.40	0.71	0.81	0.81	0.79	0.67	0.48	0.55
England	0.87	47	0.58	0.71	0.72	0.75	0.84	0.68	0.69	0.64	0.69	0.52
Finland	0.89	50	0.64	0.76	0.75	0.71	0.84	0.64	0.76	0.71	0.69	0.53
France	0.84	43	0.53	0.69	0.60	0.71	0.80	0.66	0.67	0.59	0.64	0.60
Georgia	0.72	36	0.48	0.74	0.32	0.72	0.80	0.67	0.55	0.66	0.17	0.56
Germany	0.88	49	0.57	0.75	0.73	0.72	0.85	0.68	0.68	0.70	0.72	0.56
Hong Kong SAR	0.89	50	0.58	0.72	0.60	0.80	0.84	0.76	0.72	0.74	0.65	0.63
Hungary	0.87	46	0.67	0.73	0.65	0.76	0.83	0.63	0.68	0.66	0.65	0.48
Iran, Islamic Rep. of	0.71	33	0.51	0.63	0.33	0.68	0.74	0.65	0.57	0.70	0.31	0.45
Ireland	0.87	48	0.56	0.75	0.75	0.75	0.85	0.67	0.68	0.63	0.69	0.54
Israel	0.89	51	0.59	0.73	0.58	0.80	0.85	0.74	0.77	0.67	0.70	0.65
Italy	0.86	45	0.58	0.71	0.70	0.77	0.82	0.63	0.63	0.61	0.58	0.62
Kazakhstan	0.74	36	0.58	0.66	0.14	0.67	0.77	0.62	0.66	0.63	0.63	0.44
Kuwait	0.78	36	0.57	0.64	0.42	0.69	0.74	0.68	0.66	0.57	0.47	0.47
Latvia	0.89	50	0.61	0.77	0.69	0.79	0.84	0.69	0.68	0.62	0.71	0.63
Lithuania	0.86	44	0.62	0.74	0.65	0.69	0.79	0.67	0.60	0.62	0.65	0.60
Macao SAR	0.86	45	0.55	0.66	0.54	0.79	0.83	0.70	0.67	0.67	0.59	0.63
Malta	0.84	42	0.56	0.72	0.58	0.73	0.81	0.68	0.64	0.57	0.64	0.45
Morocco	0.71	30	0.48	0.54	0.33	0.64	0.69	0.68	0.61	0.56	0.40	0.47
Netherlands	0.88	48	0.59	0.70	0.73	0.76	0.85	0.68	0.72	0.64	0.70	0.54
New Zealand	0.86	45	0.55	0.71	0.53	0.76	0.82	0.70	0.73	0.66	0.64	0.58
Northern Ireland	0.87	47	0.56	0.76	0.71	0.76	0.83	0.66	0.68	0.61	0.67	0.53
Norway (5)	0.87	46	0.63	0.73	0.73	0.75	0.84	0.63	0.69	0.63	0.62	0.50
Oman	0.78	36	0.52	0.61	0.32	0.67	0.75	0.73	0.66	0.65	0.46	0.52
Poland	0.90	53	0.63	0.77	0.70	0.73	0.86	0.71	0.72	0.66	0.75	0.69
Portugal	0.84	43	0.59	0.66	0.58	0.69	0.80	0.66	0.69	0.62	0.64	0.61
Qatar	0.83	42	0.57	0.70	0.31	0.74	0.80	0.76	0.72	0.67	0.47	0.59
Russian Federation	0.85	43	0.56	0.73	0.62	0.76	0.81	0.65	0.63	0.58	0.64	0.54
Saudi Arabia	0.77	37	0.47	0.63	0.23	0.73	0.79	0.72	0.71	0.59	0.49	0.52
Singapore	0.84	44	0.50	0.73	0.64	0.80	0.83	0.74	0.66	0.61	0.32	0.60
Slovak Republic	0.88	48	0.64	0.71	0.65	0.71	0.83	0.66	0.64	0.67	0.72	0.66
Slovenia	0.86	45	0.59	0.77	0.65	0.74	0.83	0.63	0.60	0.66	0.61	0.56
South Africa	0.75	36	0.55	0.69	0.08	0.69	0.74	0.69	0.69	0.65	0.38	0.50
Spain	0.86	45	0.56	0.66	0.63	0.74	0.82	0.69	0.69	0.62	0.62	0.64
Sweden	0.88	49	0.63	0.72	0.75	0.71	0.85	0.68	0.72	0.68	0.72	0.51
Trinidad and Tobago	0.82	39	0.58	0.65	0.49	0.75	0.81	0.64	0.68	0.60	0.48	0.50
United Arab Emirates	0.81	40	0.56	0.67	0.38	0.74	0.79	0.74	0.70	0.63	0.43	0.56
United States	0.87	46	0.59	0.71	0.61	0.77	0.82	0.68	0.71	0.64	0.65	0.59
Benchmarking Participants												
Buenos Aires, Argentina	0.87	47	0.57	0.75	0.54	0.79	0.83	0.68	0.73	0.65	0.68	0.61
Ontario, Canada	0.87	47	0.52	0.72	0.70	0.79	0.84	0.67	0.70	0.62	0.66	0.59
Quebec, Canada	0.84	42	0.52	0.68	0.69	0.75	0.82	0.60	0.61	0.60	0.65	0.54
Denmark (3)	0.83	41	0.56	0.68	0.67	0.71	0.83	0.63	0.73	0.64	0.43	0.45
Norway (4)	0.85	43	0.58	0.72	0.68	0.72	0.80	0.64	0.70	0.60	0.55	0.51
Moscow City, Russian Fed.	0.86	45	0.61	0.75	0.69	0.76	0.81	0.64	0.60	0.56	0.68	0.52
Eng/Afr/Zulu - RSA (5)	0.76	37	0.59	0.71	0.20	0.71	0.78	0.69	0.65	0.61	0.39	0.50
Andalusia, Spain	0.86	46	0.55	0.69	0.63	0.73	0.83	0.71	0.70	0.63	0.63	0.63
Madrid, Spain	0.86	46	0.57	0.67	0.68	0.73	0.82	0.70	0.68	0.61	0.64	0.64
Abu Dhabi, UAE	0.82	40	0.57	0.66	0.34	0.75	0.80	0.74	0.69	0.62	0.48	0.55
Dubai, UAE	0.83	42	0.57	0.71	0.51	0.75	0.80	0.72	0.68	0.61	0.47	0.56

*Reverse coded

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Students Like Reading* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.12	0.01	0.03
Austria	0.11	0.01	0.02
Azerbaijan	0.18	0.03	0.03
Bahrain	0.18	0.03	0.03
Belgium (Flemish)	0.14	0.02	0.02
Belgium (French)	0.11	0.01	0.01
Bulgaria	0.10	0.01	0.03
Canada	0.14	0.02	0.02
Chile	0.07	0.01	0.00
Chinese Taipei	0.20	0.04	0.03
Czech Republic	0.12	0.01	0.02
Denmark	0.23	0.05	0.04
Egypt	0.29	0.09	0.09
England	0.18	0.03	0.04
Finland	0.25	0.06	0.06
France	0.09	0.01	0.01
Georgia	0.12	0.01	0.02
Germany	0.25	0.06	0.07
Hong Kong SAR	0.19	0.04	0.04
Hungary	0.17	0.03	0.04
Iran, Islamic Rep. of	0.20	0.04	0.05
Ireland	0.18	0.03	0.04
Israel	-0.01	0.00	0.00
Italy	0.07	0.00	0.01
Kazakhstan	0.05	0.00	0.00
Kuwait	0.17	0.03	0.03
Latvia	0.06	0.00	0.01
Lithuania	0.05	0.00	0.01
Macao SAR	0.24	0.06	0.05
Malta	0.18	0.03	0.04
Morocco	0.23	0.06	0.06
Netherlands	0.20	0.04	0.04
New Zealand	0.08	0.01	0.01
Northern Ireland	0.20	0.04	0.05
Norway (5)	0.22	0.05	0.05
Oman	0.24	0.06	0.06
Poland	0.08	0.01	0.01
Portugal	0.01	0.00	0.00
Qatar	0.16	0.03	0.03
Russian Federation	0.03	0.00	0.00
Saudi Arabia	0.19	0.04	0.03
Singapore	0.22	0.05	0.05
Slovak Republic	0.16	0.02	0.02
Slovenia	0.14	0.02	0.02
South Africa	0.22	0.05	0.04
Spain	0.09	0.01	0.01
Sweden	0.21	0.04	0.05
Trinidad and Tobago	0.12	0.01	0.02
United Arab Emirates	0.19	0.04	0.04
United States	0.08	0.01	0.01
International Median	0.16	0.03	0.03
Benchmarking Participants			
Buenos Aires, Argentina	-0.09	0.01	0.01
Ontario, Canada	0.16	0.03	0.02
Quebec, Canada	0.12	0.01	0.02
Denmark (3)	0.15	0.02	0.02
Norway (4)	0.14	0.02	0.02
Moscow City, Russian Fed.	0.11	0.01	0.01
Eng/Afr/Zulu - RSA (5)	0.05	0.00	0.00
Andalusia, Spain	0.08	0.01	0.01
Madrid, Spain	0.06	0.00	0.01
Abu Dhabi, UAE	0.22	0.05	0.05
Dubai, UAE	0.12	0.01	0.02

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Students' Sense of School Belonging Scale

The *Students' Sense of School Belonging* (SSB) scale was created based on students' degree of agreement with the five statements described below.

Items in the PIRLS 2016 *Students' Sense of School Belonging* Scale

What do you think about your school? Tell how much you agree with these statements.

		Agree a lot	Agree a little	Disagree a little	Disagree a lot
ASBG12A	1) I like being in school	◯	◯	◯	◯
ASBG12B	2) I feel safe when I am at school	◯	◯	◯	◯
ASBG12C	3) I feel like I belong at this school	◯	◯	◯	◯
ASBG12D	4) Teachers at my school are fair to me	◯	◯	◯	◯
ASBG12E	5) I am proud to go to this school	◯	◯	◯	◯

High Sense of School Belonging | 9.7 | Some Sense of School Belonging | 7.3 | Little Sense of School Belonging

Item Parameters for the PIRLS 2016 *Students' Sense of School Belonging* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ASBG12A	0.35879	-0.45320	-0.71326	1.16646	1.03
ASBG12B	-0.05809	-0.47509	-0.46271	0.93780	1.02
ASBG12C	0.05363	-0.33690	-0.38143	0.71833	1.00
ASBG12D	-0.20416	-0.29014	-0.40311	0.69325	1.10
ASBG12E	-0.15017	-0.09998	-0.47379	0.57377	0.95

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Students' Sense of School Belonging* Scale

Scale Transformation Constants	
A = 7.558990	Transformed Scale Score = 7.558990 + 1.566579 • Logit Scale Score
B = 1.566579	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Students' Sense of School Belonging* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.45100	
1	4.84737	
2	5.50192	
3	5.96197	
4	6.33045	
5	6.65432	
6	6.95548	
7	7.25277	7.3
8	7.54290	
9	7.85416	
10	8.19786	
11	8.59455	
12	9.08205	
13	9.73132	9.7
14	10.70304	
15	12.65139	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Students' Sense of School Belonging* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			ASBG12A	ASBG12B	ASBG12C	ASBG12D	ASBG12E
Australia	0.82	58	0.76	0.74	0.79	0.71	0.82
Austria	0.78	53	0.74	0.78	0.70	0.65	0.76
Azerbaijan	0.72	48	0.69	0.67	0.68	0.69	0.76
Bahrain	0.77	52	0.74	0.71	0.76	0.62	0.78
Belgium (Flemish)	0.78	53	0.74	0.69	0.76	0.66	0.78
Belgium (French)	0.74	50	0.71	0.67	0.76	0.58	0.81
Bulgaria	0.76	52	0.63	0.70	0.80	0.66	0.81
Canada	0.79	54	0.72	0.73	0.75	0.68	0.81
Chile	0.82	58	0.74	0.77	0.81	0.63	0.85
Chinese Taipei	0.77	52	0.74	0.69	0.74	0.67	0.76
Czech Republic	0.72	48	0.71	0.71	0.59	0.68	0.77
Denmark	0.80	56	0.74	0.77	0.79	0.63	0.80
Egypt	0.72	47	0.65	0.67	0.70	0.66	0.75
England	0.81	57	0.75	0.73	0.76	0.69	0.83
Finland	0.81	57	0.78	0.77	0.71	0.74	0.78
France	0.70	47	0.73	0.66	0.64	0.55	0.81
Georgia	0.67	44	0.68	0.63	0.63	0.64	0.74
Germany	0.79	54	0.70	0.76	0.76	0.65	0.80
Hong Kong SAR	0.80	56	0.73	0.78	0.82	0.69	0.73
Hungary	0.78	54	0.68	0.71	0.78	0.69	0.80
Iran, Islamic Rep. of	0.63	44	0.72	0.71	0.75	0.64	0.46
Ireland	0.79	56	0.72	0.74	0.74	0.73	0.79
Israel	0.84	61	0.78	0.78	0.77	0.74	0.84
Italy	0.76	52	0.73	0.72	0.68	0.67	0.80
Kazakhstan	0.68	46	0.67	0.59	0.74	0.65	0.73
Kuwait	0.75	50	0.73	0.70	0.73	0.60	0.79
Latvia	0.77	52	0.73	0.71	0.73	0.67	0.77
Lithuania	0.73	48	0.73	0.65	0.69	0.66	0.75
Macao SAR	0.73	49	0.65	0.74	0.77	0.63	0.69
Malta	0.74	49	0.71	0.66	0.73	0.64	0.77
Morocco	0.59	39	0.52	0.67	0.72	0.52	0.67
Netherlands	0.79	55	0.76	0.73	0.75	0.70	0.78
New Zealand	0.80	56	0.72	0.74	0.77	0.69	0.81
Northern Ireland	0.78	55	0.70	0.74	0.76	0.68	0.80
Norway (5)	0.79	55	0.70	0.73	0.78	0.70	0.79
Oman	0.76	52	0.71	0.69	0.74	0.67	0.77
Poland	0.80	55	0.76	0.70	0.75	0.67	0.83
Portugal	0.74	50	0.77	0.68	0.72	0.53	0.82
Qatar	0.82	57	0.78	0.73	0.77	0.67	0.83
Russian Federation	0.72	48	0.72	0.66	0.71	0.67	0.71
Saudi Arabia	0.75	50	0.67	0.70	0.75	0.65	0.76
Singapore	0.79	55	0.76	0.72	0.77	0.65	0.80
Slovak Republic	0.78	53	0.70	0.71	0.74	0.68	0.81
Slovenia	0.78	53	0.72	0.71	0.73	0.67	0.80
South Africa	0.74	49	0.69	0.72	0.71	0.64	0.73
Spain	0.74	49	0.71	0.69	0.70	0.67	0.74
Sweden	0.82	59	0.74	0.74	0.82	0.69	0.82
Trinidad and Tobago	0.73	48	0.70	0.62	0.75	0.60	0.79
United Arab Emirates	0.79	54	0.74	0.73	0.77	0.64	0.79
United States	0.81	57	0.72	0.75	0.78	0.70	0.81
Benchmarking Participants							
Buenos Aires, Argentina	0.75	50	0.74	0.71	0.68	0.65	0.76
Ontario, Canada	0.78	54	0.72	0.72	0.75	0.66	0.82
Quebec, Canada	0.77	52	0.71	0.73	0.69	0.67	0.80
Denmark (3)	0.78	53	0.73	0.74	0.77	0.63	0.77
Norway (4)	0.78	54	0.73	0.70	0.78	0.65	0.80
Moscow City, Russian Fed.	0.77	53	0.75	0.70	0.73	0.69	0.74
Eng/Afr/Zulu - RSA (5)	0.75	50	0.69	0.70	0.78	0.61	0.75
Andalusia, Spain	0.76	51	0.72	0.68	0.73	0.66	0.78
Madrid, Spain	0.75	51	0.72	0.69	0.71	0.66	0.78
Abu Dhabi, UAE	0.78	53	0.73	0.73	0.76	0.63	0.78
Dubai, UAE	0.79	54	0.75	0.73	0.76	0.64	0.79

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Students' Sense of School Belonging* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.13	0.02	0.02
Austria	0.10	0.01	0.01
Azerbaijan	0.13	0.02	0.01
Bahrain	0.15	0.02	0.03
Belgium (Flemish)	0.16	0.03	0.03
Belgium (French)	0.09	0.01	0.01
Bulgaria	-0.07	0.00	0.00
Canada	0.14	0.02	0.02
Chile	0.17	0.03	0.03
Chinese Taipei	0.14	0.02	0.02
Czech Republic	0.06	0.00	0.01
Denmark	0.19	0.04	0.03
Egypt	-0.01	0.00	0.00
England	0.15	0.02	0.03
Finland	0.16	0.03	0.03
France	0.05	0.00	0.01
Georgia	0.05	0.00	0.01
Germany	0.18	0.03	0.03
Hong Kong SAR	0.18	0.03	0.02
Hungary	0.10	0.01	0.01
Iran, Islamic Rep. of	-0.11	0.01	0.01
Ireland	0.17	0.03	0.03
Israel	-0.04	0.00	0.00
Italy	0.10	0.01	0.01
Kazakhstan	0.03	0.00	0.00
Kuwait	0.08	0.01	0.01
Latvia	0.05	0.00	0.01
Lithuania	0.07	0.00	0.01
Macao SAR	0.14	0.02	0.02
Malta	0.18	0.03	0.04
Morocco	0.04	0.00	0.00
Netherlands	0.15	0.02	0.03
New Zealand	0.11	0.01	0.01
Northern Ireland	0.17	0.03	0.04
Norway (5)	0.15	0.02	0.02
Oman	0.16	0.02	0.03
Poland	-0.04	0.00	0.00
Portugal	0.12	0.02	0.02
Qatar	0.18	0.03	0.04
Russian Federation	0.01	0.00	0.00
Saudi Arabia	0.17	0.03	0.03
Singapore	0.11	0.01	0.01
Slovak Republic	-0.05	0.00	0.00
Slovenia	0.02	0.00	0.00
South Africa	0.08	0.01	0.01
Spain	0.08	0.01	0.01
Sweden	0.14	0.02	0.02
Trinidad and Tobago	0.12	0.02	0.02
United Arab Emirates	0.20	0.04	0.05
United States	0.17	0.03	0.03
International Median	0.12	0.02	0.01
Benchmarking Participants			
Buenos Aires, Argentina	0.01	0.00	0.00
Ontario, Canada	0.15	0.02	0.02
Quebec, Canada	0.11	0.01	0.01
Denmark (3)	0.16	0.03	0.02
Norway (4)	0.11	0.01	0.01
Moscow City, Russian Fed.	0.08	0.01	0.01
Eng/Afr/Zulu - RSA (5)	-0.07	0.00	0.00
Andalusia, Spain	0.05	0.00	0.01
Madrid, Spain	0.05	0.00	0.01
Abu Dhabi, UAE	0.15	0.02	0.03
Dubai, UAE	0.20	0.04	0.05

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Teacher Job Satisfaction Scale

The *Teacher Job Satisfaction* (TJS) scale was created based on the degree that teachers responded positively to the five items below.

Items in the PIRLS 2016 *Teacher Job Satisfaction* Scale[1]

How often do you feel the following way about being a teacher?

		Very often	Often	Sometimes	Never or almost never
ATBG10A	1) I am content with my profession as a teacher	◯	◯	◯	◯
ATBG10B	2) I find my work full of meaning and purpose	◯	◯	◯	◯
ATBG10C	3) I am enthusiastic about my job	◯	◯	◯	◯
ATBG10D	4) My work inspires me	◯	◯	◯	◯
ATBG10E	5) I am proud of the work I do	◯	◯	◯	◯

Very Satisfied | 10.2 | **Somewhat Satisfied** | 6.2 | **Less than Satisfied**

1 For the purpose of scaling, categories in which there were very few respondents were combined. The categories "Sometimes" and "Never or almost never" were combined for all variables. The scale statistics that are reported herein reflect analysis of the items following collapsing.

Item Parameters for the PIRLS 2016 *Teacher Job Satisfaction* Scale

Item	delta	tau_1	tau_2	Infit
ATBG10A	0.37535	-2.26426	2.26426	1.11
ATBG10B	-0.78996	-2.26314	2.26314	1.05
ATBG10C	0.06308	-2.28012	2.28012	0.89
ATBG10D	0.46060	-2.32478	2.32478	0.91
ATBG10E	-0.10907	-1.95730	1.95730	1.14

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the PIRLS 2016 *Teacher Job Satisfaction* Scale

Scale Transformation Constants	
A = 8.192812	Transformed Scale Score = 8.192812 + 0.804273 • Logit Scale Score
B = 0.804273	

Equivalence Table of the Raw Score and Transformed Scale Scores for the PIRLS 2016 *Teacher Job Satisfaction* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	4.38704	
1	5.47445	
2	6.12206	6.2
3	6.68375	
4	7.29255	
5	8.17713	
6	9.07681	
7	9.70719	
8	10.27783	10.2
9	10.92662	
10	11.99778	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the PIRLS 2016 *Teacher Job Satisfaction* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item				
			ATBG10A	ATBG10B	ATBG10C	ATBG10D	ATBG10E
Australia	0.90	71	0.87	0.82	0.89	0.87	0.74
Austria	0.86	65	0.76	0.75	0.85	0.86	0.80
Azerbaijan	0.79	55	0.71	0.68	0.73	0.81	0.77
Bahrain	0.88	69	0.84	0.80	0.85	0.83	0.82
Belgium (Flemish)	0.89	69	0.81	0.82	0.85	0.86	0.81
Belgium (French)	0.90	73	0.88	0.87	0.90	0.88	0.73
Bulgaria	0.87	66	0.75	0.82	0.85	0.85	0.79
Canada	0.89	70	0.81	0.84	0.87	0.88	0.78
Chile	0.86	65	0.79	0.82	0.82	0.86	0.74
Chinese Taipei	0.94	80	0.91	0.88	0.93	0.89	0.86
Czech Republic	0.91	73	0.82	0.88	0.87	0.89	0.82
Denmark	0.89	69	0.81	0.80	0.89	0.85	0.79
Egypt	0.74	50	0.80	0.66	0.73	0.74	0.60
England	0.90	72	0.80	0.87	0.89	0.90	0.78
Finland	0.92	76	0.86	0.83	0.91	0.90	0.84
France	0.84	62	0.77	0.63	0.89	0.88	0.74
Georgia	0.84	61	0.73	0.68	0.84	0.82	0.82
Germany	0.87	66	0.81	0.75	0.84	0.87	0.78
Hong Kong SAR	0.94	81	0.90	0.92	0.91	0.87	0.89
Hungary	0.89	69	0.80	0.83	0.87	0.85	0.80
Iran, Islamic Rep. of	0.72	52	0.76	0.69	0.81	0.84	0.45
Ireland	0.90	72	0.80	0.87	0.86	0.89	0.82
Israel	0.91	74	0.84	0.85	0.89	0.88	0.84
Italy	0.89	69	0.78	0.77	0.89	0.86	0.85
Kazakhstan	0.87	66	0.77	0.80	0.86	0.81	0.81
Kuwait	0.81	59	0.74	0.75	0.85	0.70	0.78
Latvia	0.85	62	0.76	0.76	0.75	0.84	0.82
Lithuania	0.87	65	0.81	0.65	0.85	0.87	0.83
Macao SAR	0.91	74	0.88	0.89	0.87	0.82	0.86
Malta	0.88	68	0.80	0.85	0.83	0.85	0.79
Morocco	0.85	63	0.77	0.75	0.85	0.82	0.78
Netherlands	0.87	66	0.75	0.77	0.86	0.85	0.84
New Zealand	0.88	69	0.78	0.81	0.88	0.90	0.77
Northern Ireland	0.89	70	0.82	0.87	0.86	0.89	0.73
Norway (5)	0.89	70	0.74	0.80	0.88	0.88	0.87
Oman	0.82	60	0.77	0.71	0.87	0.83	0.67
Poland	0.92	77	0.89	0.88	0.88	0.88	0.87
Portugal	0.88	67	0.82	0.81	0.87	0.83	0.78
Qatar	0.88	68	0.73	0.84	0.86	0.88	0.80
Russian Federation	0.85	63	0.79	0.72	0.81	0.83	0.80
Saudi Arabia	0.86	65	0.88	0.78	0.76	0.78	0.84
Singapore	0.96	85	0.91	0.92	0.94	0.94	0.92
Slovak Republic	0.90	71	0.81	0.80	0.88	0.87	0.85
Slovenia	0.90	70	0.82	0.78	0.87	0.87	0.84
South Africa	0.90	71	0.79	0.87	0.85	0.87	0.83
Spain	0.85	62	0.79	0.75	0.83	0.85	0.71
Sweden	0.86	64	0.81	0.69	0.87	0.85	0.78
Trinidad and Tobago	0.91	74	0.83	0.87	0.92	0.88	0.80
United Arab Emirates	0.87	66	0.81	0.77	0.85	0.85	0.77
United States	0.90	72	0.84	0.89	0.87	0.87	0.77
Benchmarking Participants							
Buenos Aires, Argentina	0.75	50	0.74	0.71	0.78	0.69	0.63
Ontario, Canada	0.90	71	0.79	0.86	0.89	0.90	0.76
Quebec, Canada	0.87	66	0.75	0.71	0.88	0.88	0.81
Denmark (3)	0.89	70	0.83	0.80	0.87	0.88	0.79
Norway (4)	0.91	74	0.84	0.84	0.90	0.90	0.82
Moscow City, Russian Fed.	0.85	62	0.78	0.77	0.79	0.85	0.75
Eng/Afr/Zulu - RSA (5)	0.91	74	0.74	0.90	0.92	0.90	0.81
Andalusia, Spain	0.83	61	0.78	0.78	0.80	0.87	0.65
Madrid, Spain	0.88	68	0.84	0.84	0.77	0.83	0.85
Abu Dhabi, UAE	0.89	69	0.85	0.80	0.88	0.86	0.77
Dubai, UAE	0.89	69	0.81	0.81	0.88	0.87	0.77

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the PIRLS 2016 *Teacher Job Satisfaction* Scale and PIRLS 2016 Reading Achievement

Country	Pearson's Correlation with Reading Achievement		Variance in Reading Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Australia	0.05	0.00	0.00
Austria	0.00	0.00	0.00
Azerbaijan	0.06	0.00	0.00
Bahrain	0.08	0.01	0.00
Belgium (Flemish)	0.01	0.00	0.00
Belgium (French)	0.09	0.01	0.01
Bulgaria	0.05	0.00	0.00
Canada	0.00	0.00	0.00
Chile	-0.01	0.00	0.00
Chinese Taipei	-0.01	0.00	0.00
Czech Republic	0.03	0.00	0.00
Denmark	0.02	0.00	0.00
Egypt	0.00	0.00	0.00
England	0.00	0.00	0.00
Finland	-0.03	0.00	0.00
France	0.02	0.00	0.00
Georgia	0.04	0.00	0.00
Germany	0.10	0.01	0.01
Hong Kong SAR	0.01	0.00	0.00
Hungary	0.04	0.00	0.00
Iran, Islamic Rep. of	0.06	0.00	0.00
Ireland	0.03	0.00	0.00
Israel	-0.04	0.00	0.01
Italy	-0.01	0.00	0.00
Kazakhstan	0.08	0.01	0.00
Kuwait	-0.07	0.01	0.01
Latvia	0.09	0.01	0.01
Lithuania	0.10	0.01	0.01
Macao SAR	0.09	0.01	0.01
Malta	0.05	0.00	0.01
Morocco	0.16	0.03	0.02
Netherlands	-0.02	0.00	0.01
New Zealand	0.08	0.01	0.00
Northern Ireland	-0.01	0.00	0.00
Norway (5)	0.04	0.00	0.00
Oman	0.01	0.00	0.00
Poland	-0.01	0.00	0.00
Portugal	0.05	0.00	0.00
Qatar	-0.08	0.01	0.00
Russian Federation	0.02	0.00	0.00
Saudi Arabia	0.15	0.02	0.02
Singapore	-0.02	0.00	0.00
Slovak Republic	0.01	0.00	0.00
Slovenia	-0.01	0.00	0.00
South Africa	-0.03	0.00	0.03
Spain	0.07	0.01	0.00
Sweden	0.01	0.00	0.00
Trinidad and Tobago	0.03	0.00	0.00
United Arab Emirates	-0.03	0.00	0.00
United States	0.09	0.01	0.01
International Median	0.02	0.00	0.00
Benchmarking Participants			
Buenos Aires, Argentina	0.06	0.00	0.00
Ontario, Canada	0.02	0.00	0.00
Quebec, Canada	-0.06	0.00	0.00
Denmark (3)	0.01	0.00	0.00
Norway (4)	0.03	0.00	0.00
Moscow City, Russian Fed.	-0.02	0.00	0.00
Eng/Afr/Zulu - RSA (5)	-0.15	0.02	0.02
Andalusia, Spain	0.02	0.00	0.00
Madrid, Spain	0.11	0.01	0.01
Abu Dhabi, UAE	-0.08	0.01	0.01
Dubai, UAE	-0.02	0.00	0.00

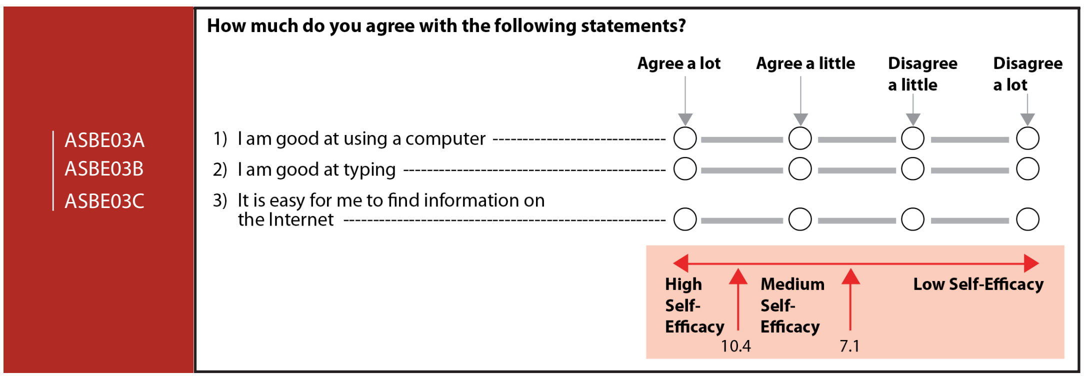
SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Appendix 14B: ePIRLS 2016 Context Questionnaire Scale

Self-Efficacy for Computer Use Scale

The *Self-Efficacy for Computer Use* (SEC) scale was created based on students' degree of agreement with the three statements described below.

Items in the ePIRLS 2016 *Self-Efficacy for Computer Use* Scale



Item Parameters for the ePIRLS 2016 *Self-Efficacy for Computer Use* Scale

Item	delta	tau_1	tau_2	tau_3	Infit
ASBE03A	-0.30146	-0.63034	-0.70712	1.33746	0.95
ASBE03B	0.28602	-0.73511	-0.61080	1.34591	0.97
ASBE03C	0.01544	-0.88525	-0.38788	1.27313	1.04

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Scale Transformation Constants for the ePIRLS 2016 *Self-Efficacy for Computer Use* Scale

Scale Transformation Constants	
A = 7.582113	Transformed Scale Score = 7.582113 + 1.557903 • Logit Scale Score
B = 1.557903	

Equivalence Table of the Raw Score and Transformed Scale Scores for the ePIRLS 2016 *Self-Efficacy for Computer Use* Scale

Raw Score	Transformed Scale Score	Cutpoint
0	3.53502	
1	5.02075	
2	5.81380	
3	6.43388	
4	7.02434	7.1
5	7.63722	
6	8.34823	
7	9.25796	
8	10.49854	10.4
9	12.62271	

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Cronbach's Alpha Reliability Coefficient and Principal Components Analysis of the Items in the ePIRLS 2016 *Self-Efficacy for Computer Use* Scale

Country	Cronbach's Alpha Reliability Coefficient	Percent of Variance Explained	Component Loadings for Each Item		
			ASBE03A	ASBE03B	ASBE03C
Canada	0.57	54	0.78	0.77	0.65
Chinese Taipei	0.66	60	0.77	0.79	0.76
Denmark	0.68	61	0.79	0.81	0.75
Georgia	0.51	52	0.78	0.61	0.77
Ireland	0.67	61	0.82	0.81	0.70
Israel	0.63	58	0.79	0.79	0.71
Italy	0.69	63	0.82	0.80	0.75
Norway (5)	0.66	60	0.81	0.79	0.72
Portugal	0.66	60	0.84	0.81	0.67
Singapore	0.64	59	0.81	0.78	0.69
Slovenia	0.68	62	0.83	0.79	0.74
Sweden	0.71	64	0.85	0.83	0.71
United Arab Emirates	0.57	54	0.76	0.75	0.69
United States	0.58	55	0.78	0.75	0.69
Benchmarking Participants					
Abu Dhabi, UAE	0.55	53	0.75	0.76	0.67
Dubai, UAE	0.58	54	0.76	0.77	0.68

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

Relationship Between the ePIRLS 2016 *Self-Efficacy for Computer Use* Scale and ePIRLS 2016 Online Informational Reading Achievement

Country	Pearson's Correlation with ePIRLS Achievement		Variance in ePIRLS Achievement Accounted for by Difference Between Regions of the Scale (η^2)
	(r)	(r^2)	
Canada	0.08	0.01	0.01
Chinese Taipei	0.08	0.01	0.01
Denmark	0.14	0.02	0.03
Georgia	0.14	0.02	0.02
Ireland	0.06	0.00	0.01
Israel	0.12	0.01	0.02
Italy	0.03	0.00	0.00
Norway (5)	0.12	0.01	0.02
Portugal	0.10	0.01	0.01
Singapore	0.10	0.01	0.01
Slovenia	0.06	0.00	0.00
Sweden	0.07	0.01	0.02
United Arab Emirates	0.16	0.03	0.03
United States	0.07	0.00	0.01
International Median	0.09	0.01	0.01
Benchmarking Participants			
Abu Dhabi, UAE	0.17	0.03	0.04
Dubai, UAE	0.14	0.02	0.02

SOURCE: IEA's Progress in International Reading Literacy Study – PIRLS 2016

BOSTON
COLLEGE

pirls.bc.edu

ISBN: 978-1-889938-44-8

IEA

TIMSS & PIRLS
International Study Center
Lynch School of Education
BOSTON COLLEGE

© IEA, 2017
International Association
for the Evaluation of
Educational Achievement