CHAPTER 1

Developing the PIRLS 2021 Achievement Instruments

Erin Wry
Ina V.S. Mullis

Unique Characteristics of PIRLS 2021

The PIRLS assessment instruments are designed to provide valid measures of students' reading abilities in their fourth year of schooling. The general approach to developing the PIRLS achievement materials is similar from one assessment cycle to the next, but each cycle has some unique characteristics that influence instrument development. In addition to continuing the PIRLS trend lines and monitoring changes in educational achievement, PIRLS 2021 was remarkable for several reasons:

- PIRLS 2021 marked an important milestone in the transition to a fully digital assessment, with about half of the participating countries administering PIRLS digitally. The digital administration of PIRLS offers an engaging, interactive, and visually attractive assessment that included digital versions of the PIRLS literary and informational texts as well as ePIRLS online reading tasks.
- As part of the transition to a digital assessment, a digital administration system was developed that included a new innovative user interface for administering informational and literary texts and questions. The digital administration system also allowed for the integration of ePIRLS and overall helped increase operational efficiency in assessment delivery, data entry, and scoring.
- Building on past efforts to accommodate a wider range of student reading abilities, PIRLS 2021 implemented a new group adaptive assessment design. By capitalizing on the range in difficulty of the texts developed for PIRLS and PIRLS Literacy in 2016, and continuing to expand the difficulty range with the newly developed texts, PIRLS 2021 offered one unified assessment that can better measure the range of high, medium, and low reading abilities found in each of the PIRLS participating countries.

https://doi.org/10.6017/lse.tpisc.tr2101.kb7549

Transitioning to a Digital Assessment

The transition to a digital assessment enables PIRLS to continue to reflect the current practices and authentic reading situations of young students in an increasingly technological world. However, because not all PIRLS countries were prepared to conduct a fully digital assessment in 2021, IEA decided to implement the transition over two assessment cycles—PIRLS 2021 and PIRLS 2026. In PIRLS 2021, about half of the participating countries elected to administer PIRLS digitally (digitalPIRLS), while the rest of the countries administered PIRLS in paper-and-pencil format as in previous assessment cycles (paperPIRLS), with the eventual goal of all countries participating digitally in PIRLS 2026.

To ensure comparability across formats, the digitalPIRLS assessment contained the same literary and informational texts and items administered in paperPIRLS, but presented in the digital interface. In addition, the digitalPIRLS assessment design also included five ePIRLS online informational tasks—three from ePIRLS 2016 and two newly developed for PIRLS 2021. For countries participating in digitalPIRLS, the five tasks were integrated into the rotation of literary and informational blocks and administered to students via the player software.

The digital assessment offers several important operational advantages. Without the need for printing and tracking materials manually, the digital software can accommodate complex designs that can be updated and adjusted more efficiently. Furthermore, the digital entry of the student answers eliminates the need for manual data entry and also allows for many item types to be machine scored, reducing the resources and time needed for scoring procedures. In addition to the operational benefits, the digital assessment software has the capacity to collect process data on students' interactions with the test, such as students' time spent on items, use of interface tools, and navigation approaches. This process data can be used in further analyses to better understand students' test taking and reading behaviors.

PIRLS 2021 Group Adaptive Design

As evidenced throughout PIRLS' 20-year trend measurement, achievement levels in reading comprehension vary widely across and within countries. This makes it a challenge to develop an assessment that matches the reading ability of all participating students. If the assessment is too difficult or too easy for a population, there is less information available to accurately measure performance. To address this challenge in past cycles, PIRLS offered less difficult versions of its assessment materials, beginning with prePIRLS in 2011 followed by PIRLS Literacy in 2016. These efforts were successful in expanding PIRLS coverage of students at the lower end of the ability distribution, but required additional operational and methodological efforts. Moreover, these less difficult versions did not address the need for more challenging material for higher achieving students.

PIRLS 2021 reflects a broader range of assessment difficulty and better targeting of student ability through a new group adaptive assessment design (see rationale underlying adaptive assessment designs). The PIRLS 2021 design comprises three levels of passage difficulty—difficult, medium, and easy—that are combined into two levels of booklet difficulty. More difficult booklets are composed of two difficult passages or one medium and one difficult passage, while less difficult booklets consist of an easy and a medium passage or two easy passages. Each country still administers the entire assessment, but the balance of more difficult and less difficult booklets varies with the reading achievement level of the students in the country. This approach enables a better match between assessment difficulty and student ability in each country's population by having a greater proportion of more difficult booklets in countries with relatively high achievement and a greater proportion of less difficult booklets in countries with relatively low achievement. Accordingly, the new PIRLS 2021 group adaptive design maximizes the information obtained from the assessment while minimally changing existing procedures and time requirements.

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 test also needs to be updated with each cycle to prevent the assessment from becoming dated and no longer relevant to current learning goals and policy issues. It is important that PIRLS reflects the most recent advancements 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, PIRLS has a specific design for rotating texts and items out of the assessment after each cycle and replacing them with newly developed texts and items. In this approach, each PIRLS assessment includes texts 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 texts and items are made available after each cycle on a restricted use basis for educational and research purposes (please see https://www.iea.nl/copyrightnotice).

Overview of the PIRLS 2021 Development Process

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 online texts,

and subsequently the items, must be reviewed by experts and agreed upon by the diverse group of participating countries.

The TIMSS & PIRLS International Study Center at Boston College uses a collaborative process to select the passages and develop the new texts and 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/online texts
- 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/tasks 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 2021, Executive Directors Ina Mullis and Michael Martin managed the assessment development process, and Erin Wry served as the PIRLS Coordinator.

Also playing a key role in achievement item development were the National Research Coordinators (NRCs) who are 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 texts and develop new test items.

The TIMSS & PIRLS International Study Center prepared an international version of all the PIRLS 2021 achievement texts and items in English. Subsequently, the materials were translated by participating countries into their languages of instruction with the goal of creating high quality translations that were appropriately adapted for the national context, and at the same time, remained internationally comparable. Therefore, a significant portion of the NRCs' development and review efforts were dedicated to ensuring that the achievement texts and items could be translated accurately.

Additional advice was provided through periodic reviews by the PIRLS 2021 Reading Development Group (RDG). As international experts in assessing reading comprehension, RDG members for each PIRLS cycle are nominated by participating countries and provide guidance in developing the PIRLS achievement materials. Exhibit 1.1 lists the ten members of the PIRLS 2021 RDG.

Exhibit 1.1: PIRLS 2021 Reading Development Group (RDG)

Julian Fraillon
IEA Amsterdam
Australia

Jan Mejding
Aarhus University
Department of Education (DPU)
Denmark

Liz Twist
National Foundation for Educational Research
England

Galina Zuckerman
Psychological Institute
Russian Academy of Education
Russian Federation

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

Latifa Alfalasi
Ministry of Education
United Arab Emirates

Verónica Díez Girado
National Institute for Educational Assessment
Ministry of Education
Spain

Karen Wixson
Educational Testing Service
United States

Elizabeth Pang
Curriculum Planning and Development Division
Ministry of Education
Singapore

Jenny Wiksten Folkeryd
Uppsala University
Department of Education
Sweden

RDG members met four times for PIRLS 2021. At the first RDG meeting in Rotterdam, The Netherlands (May 2018), the RDG reviewed the reading frameworks, potential passages, ePIRLS prototypes, and draft item writing guidelines. At the second meeting in Dubai, UAE (April 2019), the RDG reviewed PIRLS field test passages and items and ePIRLS field test tasks and items. At the third meeting which took place virtually due to the COVID-19 Pandemic (July 2020), the RDG reviewed field test results and made recommendations to the NRCs regarding which passages and items to include in the 2021 assessment. At the fourth and final meeting which also took place virtually (July 2022), the RDG conducted the PIRLS 2021 scale anchoring process.

Between RDG meetings, a subgroup of several RDG members served as a "working group" to assist in completing specific tasks, such as reviewing passages and developing items and scoring guides.

The PIRLS 2021 Development Schedule

The first year of the PIRLS 2021 assessment cycle was devoted to updating the assessment framework, as well as piloting and refining the new digitalPIRLS interface, and identifying appropriate texts. The second year of PIRLS 2021 was a continuation of these efforts and included the additional task of developing new achievement items based on the proposed texts. The third year of the cycle was devoted to selecting the text and items for the PIRLS 2021 Field Test and collecting field test data. Typically, the PIRLS field test is conducted from March to May in the third year of the assessment cycle. However due to the unprecedented circumstances brought on by the COVID-19 pandemic in the spring of 2020, many countries had to alter or delay their field test plans. Despite these challenging circumstances, more than half of the countries were still able to collect field test data, which provided adequate data to evaluate the passages and items for inclusion in PIRLS 2021 (see following section about field test).

Exhibit 1.2 shows the PIRLS 2021 development schedule from updating the frameworks to data collection.

Exhibit 1.2: PIRLS 2021 Development Schedule for Achievement Instruments

Date(s)		Group and Activity
September	2017	TIMSS & PIRLS International Study Center began updating the PIRLS Assessment Framework for PIRLS 2021 focusing on the curricular emphases in reading described in the PIRLS 2016 Encyclopedia
September–October	2017	Reading Development Group (RDG) conducted an online review of the draft PIRLS 2021 Reading Assessment Framework and provided feedback to the TIMSS & PIRLS International Study Center. The staff then met with the working group consultants to incorporate the RDG's comments (Boston, USA)
January	2018	TIMSS & PIRLS International Study Center began designing the digitalPIRLS user interface and digital item types
February	2018	NRCs reviewed and provided feedback on the draft PIRLS 2021 Reading Assessment Framework and the TIMSS & PIRLS International Study Center presented passages selection criteria and 9 potential passages for PIRLS 2021 (1st NRC Meeting—Hamburg Germany)
February–July	2018	TIMSS & PIRLS International Study Center created an online discussion board to post potential texts submitted by NRCs, the RDG, and consultants for consideration for the PIRLS 2021 field test. Staff also revised the draft PIRLS 2021 Reading Assessment Framework based on feedback from NRCs
April	2018	TIMSS & PIRLS International Study Center provided demo versions of the digitalPIRLS user interface to the United States and New Zealand to try out in fourth grade classrooms in preparation of cognitive labs
May	2018	Reading Development Group (RDG) reviewed the updated Reading Assessment Framework, proposed field test passages/tasks and text maps, the new digitalPIRLS interface design, and the draft PIRLS 2021 Item Writing Guidelines (1st RDG Meeting—Rotterdam, The Netherlands)

Exhibit 1.2: PIRLS 2021 Development Schedule for Achievement Instruments (Continued)

Date(s)		Group and Activity
June	2018	The TIMSS & PIRLS International Study Center continued to refine the digitalPIRLS interface based off of feedback from the United States, New Zealand, and the RDG, and developed directions and discussion protocol for the AIR user interface cognitive labs
July	2018	American Institutes for Research (AIR) conducted cognitive laboratories for two trend PIRLS passages presented in the proposed digitalPIRLS interface which were observed by staff from the TIMSS & PIRLS International Study Center. Following the labs, AIR provided a comprehensive report of student feedback that was used to further improve and refine the digitalPIRLS interface
July–September	2018	Staff at the TIMSS & PIRLS International Study Center reviewed feedback from the online discussion board – 34 passages were posted to the discussion board, 16 of which received favorable reviews from the NRCs
October	2018	RDG members participated in an online review of 16 potential passages for the PIRLS 2021 Field Test and submitted feedback to the TIMSS & PIRLS International Study Center
October	2018	Staff at the TIMSS & PIRLS International Study Center and working group consultants compiled feedback from the RDG to identify the most appropriate passages to propose for the PIRLS 2021 field test and continued refining the two newly developed ePIRLS tasks (Boston, USA)
December	2018	NRCs performed final review of the *PIRLS 2021 Assessment Framework*, selected passages and ePIRLS tasks for the field test, and developed draft field test items using the PIRLS 2021 Item Writing Guidelines (2nd NRC Meeting—Taipei City, Chinese Taipei)
April	2019	TIMSS & PIRLS International Study Center published the PIRLS 2021 Assessment Framework
April	2019	RDG reviewed draft field test passages/tasks and item sets (2nd RDG Meeting—Dubai, UAE)
June	2019	NRCs reviewed and approved PIRLS passages and ePIRLS tasks for the PIRLS 2021 Field Test (3rd NRC Meeting—Rome, Italy)
June–July	2019	TIMSS & PIRLS International Study Center assembled all PIRLS 2021 Field Test instruments (paper and digital) and released international instruments to countries to begin translating along with documentation on preparing national instruments and conducting the field test procedures
October	2019	TIMSS & PIRLS International Study Center began adapting the 12 paper-based trend passages to a digital format, in preparation for data collection
October	2019	TIMSS & PIRLS International Study Center worked with seven English-speaking countries to collect student responses to newly developed PIRLS constructed response items in order to develop scoring training materials for the field test
November	2019	TIMSS & PIRLS International Study Center and working group consultants reviewed students' responses from the pilot and developed scoring training materials for the 4th NRC Meeting (Boston, USA)
March	2020	NRCs received scoring training for PIRLS constructed response field test items (4th NRC Meeting—Belgrade, Serbia)
March–June	2020	Countries conducted the PIRLS 2021 Field Test
May–June	2020	Countries submitted PIRLS 2021 Field Test achievement data for analysis and review

Exhibit 1.2: PIRLS 2021 Development Schedule for Achievement Instruments (Continued)

Date(s)		Group and Activity
June	2020	IEA Hamburg completed data processing and the TIMSS & PIRLS International Study Center reviewed field test item statistics
June–July	2020	TIMSS & PIRLS International Study Center assembled proposed passages and items for data collection in preparation for the review by the Reading Development Group and NRCs
July	2020	RDG reviewed proposed passages/tasks and items in conjunction with field test results (3rd RDG Meeting—virtual)
August	2020	NRCs reviewed and approved the PIRLS passages, ePIRLS tasks, and items proposed for PIRLS 2021 data collection (5th NRC Meeting—virtual)
August	2020	TIMSS & PIRLS International Study Center finalized all PIRLS 2021 Data Collection instruments and released the international instruments to countries for translation
September–October	2020	Operational Field Tests
October–December	2020	Southern Hemisphere countries conducted PIRLS 2021 Data Collection
November	2020	NRCs from Southern Hemisphere countries received scoring training for PIRLS 2021 constructed response items (virtual)
November	2020	TIMSS & PIRLS International Study Center finalized scoring guides and training materials for the PIRLS 2021 constructed response items and distributed them to NRCs
March	2021	NRCs from Northern Hemisphere countries received scoring training for PIRLS 2021 constructed response items (6th NRC Meeting—virtual)
March–June	2021	Northern Hemisphere countries conducted PIRLS 2021 Data Collection
August–December	2021	Northern Hemisphere countries that delayed the assessment to the beginning of fifth grade conducted PIRLS 2021 Data Collection; Southern Hemisphere countries that assessed students at the end of the fourth grade but one year later conducted PIRLS 2021 Data Collection
March–July	2022	Northern Hemisphere countries that delayed the assessment by one year conducted PIRLS 2021 Data Collection

Updating the Reading Assessment Framework for PIRLS 2021

As the first step in developing PIRLS 2021, staff at the TIMSS & PIRLS International Study Center, in collaboration with reading experts, reviewed and updated the PIRLS Reading Assessment Framework. Given the overarching goal of measuring trends, the assessment framework cannot be changed drastically across cycles, however, it is updated to introduce fresh ideas and current information about curricula, standards, and instruction in reading. The first chapter of the *PIRLS 2021 Assessment Frameworks* contains the PIRLS 2021 Reading Assessment Framework (Mullis & Martin, 2019).

Through an iterative process, the updated Reading Assessment Framework was first reviewed by the RDG, and then revised again in preparation for the NRC's review. At the 1st PIRLS 2021 NRC Meeting in February 2018 (Hamburg, Germany), NRCs reviewed the proposed framework and consulted with their national experts about the updates for PIRLS 2021. Following the discussion at the 1st NRC meeting, the framework was updated again based on country feedback and then reviewed by both the RDG and NRCs a final time prior to publication in April 2019.

Developing the digitalPIRLS User Interface (UI)

In parallel with updating the PIRLS 2021 assessment framework, the TIMSS & PIRLS International Study Center began designing a new digital interface for administering literary and informational texts and items in PIRLS 2021. Throughout its development, the digitalPIRLS UI underwent a series of external reviews, beginning with the 1st PIRLS 2021 NRC Meeting in February 2018 (Hamburg, Germany). After implementing feedback from the NRC review, the TIMSS & PIRLS International Study Center provided two English speaking countries with demonstration versions of the UI to try out in fourth-grade classrooms in preparation of the cognitive labs.

In July of 2018, the TIMSS & PIRLS International Study Center arranged for the American Institute for Research (AIR) to conduct cognitive labs with fourth-grade students in Washington, D.C., which were observed by staff at the TIMSS & PIRLS International Study Center. The focus of the cognitive labs was to investigate the student experience with the new assessment interface design using two released PIRLS passages—*Macy and the Red Hen* (literary passage) and *The Green Sea Turtle's Journey of a Lifetime* (informational passage)—presented in the digitalPIRLS interface.

Interview protocols were developed for the labs that involved both a think aloud aspect and retrospective aspect for one-on-one interviews with the students. During the cognitive labs, nine students, representing a range of reading abilities, computer efficacies, and socioeconomic backgrounds were administered the assessment and told to voice any thoughts, questions, or frustrations with the interface throughout the test. As the students worked, their interactions with the software and verbal and nonverbal feedback were monitored and recorded. During the post-assessment interviews, particular attention was dedicated to the students' experiences with using the scrollbar to navigate between pages, opening and navigating through the question window, using a mouse and/or keyboard to answer questions, and activating and deactivating the highlighter to highlight text. Following the cognitive labs, AIR provided the TIMSS & PIRLS International Study Center with a comprehensive report of the students' experiences and feedback that was used to further revise the interface for the PIRLS 2021 field test and data collection.

Identifying Texts for PIRLS and ePIRLS

In total, 18 new passages and item sets, as well as 2 new ePIRLS tasks were developed in consideration for the 12 passages and 2 ePIRLS tasks needed for the PIRLS 2021 field test. Identifying appropriate texts for PIRLS and ePIRLS is critical 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. In order to reflect the goal of approximating an authentic reading experience in the assessment, the texts included in PIRLS—whether presented digitally, in an online format, or in print—need to be typical of those read by students in their everyday experiences in and outside of school.

At the beginning of the assessment cycle, the TIMSS & PIRLS International Study Center called upon NRCs to propose potential literary and informational texts for PIRLS 2021. The criteria for suitable texts was discussed at the first NRC meeting in Hamburg in February 2018. RDG member, Liz Twist from NFER, explained that literary and informational 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 2018, the TIMSS & PIRLS International Study Center created a discussion board for NRCs to review submitted texts. 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 and PIRLS International Study Center computed the word count and several readability indices for each passage as a quantitative check of the grade appropriateness of potential texts. The Flesch-Kincaid Grade Level Formula (Kindcaid et al., 1975), known for its suitability for a wide range of texts and its extensive use in education, was used as the primary index of readability. Two other measures of readability, the Lexile Level and the ETS Text Evaluator, were also calculated. This quantitative information was provided alongside the texts on the discussion board for NRCs to consider during their reviews.

All of the proposed texts underwent extensive review by the RDG and the NRCs. In particular, considerable effort was expended to ensure that the texts and websites had the following characteristics:

- Clarity and coherence
- Appropriate content across countries and cultures
- Interesting, engaging content for a wide range of students
- Adequate basis for assessing the full range of comprehension processes

The availability of assessment resources places some limits on the length of texts, because students need time to read the entire passage and answer comprehension questions. Consistent with the range in difficulty across PIRLS, the length of literary and informational texts generally average from about 500 to 800 words.

The texts included in PIRLS 2021 were written for a young audience and have an appropriate level of linguistic features and density of information. In the context of an international study, attaining authenticity in assessment reading experience may be somewhat constrained by the need to translate the texts into numerous languages. Therefore, a significant portion of the NRC's development and review effort is dedicated to ensuring that the chosen texts can be translated without loss of clarity in meaning, or in potential for student engagement. 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.

In addition to the literary and informational texts, 2 new ePIRLS online informational reading tasks were developed for PIRLS 2021. 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 PIRLS passages, taking into consideration suitability for fourth grade students regarding content, interest, reading ability, complexity, and cultural sensitivity. However, reflecting the fact that online reading often involves sorting through more information than may be necessary, each ePIRLS task involves approximately three different websites with about five to ten web pages, averaging about 1000 words in total.

Writing and Reviewing the PIRLS 2021 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. The 2nd PIRLS NRC meeting in Taipei City, Chinese Taipei in December 2018, was devoted to determining the texts that would be field tested followed by a workshop for developing the field test items and draft scoring guides.

Prior to the PIRLS item writing workshop, staff at the TIMSS & PIRLS International Study Center 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 2021 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. The PIRLS 2021 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 2021 items.

At the 2nd NRC Meeting, NRCs voted for 12 of the 18 proposed passages and 2 ePIRLS tasks to be include in the field test. Once the field test materials were selected, NRCs, together with experienced item writers from participating countries and staff from the TIMSS & PIRLS International Study Center, developed items for the proposed PIRLS passages and ePIRLS tasks. During the PIRLS 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 2021 framework. The TIMSS & PIRLS International Study Center staff 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. Collectively, 607 items for PIRLS texts and 22 items for ePIRLS tasks were developed.

Exhibit 1.3 shows the number of participants in the PIRLS 2021 item writing workshop and the number of items written.

Exhibit 1.3: PIRLS 2021 Item Writing Workshop to Develop Field Test Items

Attendees	
Number of Countries and Benchmarking Entities	45
Number of Country Representatives	126
Approximate Number of Field Test Items Written at Item Writing Workshop	
PIRLS	607
ePIRLS	22

Following the item writing workshop, the TIMSS & PIRLS International Study Center and members of the PIRLS working group thoroughly reviewed the draft set of texts and field test items. This was followed by a review by RDG members at the 2nd PIRLS 2021 RDG Meeting in Dubai, UAE, and then again by the NRCs at the 3rd PIRLS 2021 NRC meeting in Rome, Italy. Once all reviews were complete, 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 instruments to the NRCs to begin translating the field test materials into their languages of instruction.

The PIRLS 2021 Field Test

In preparation for data collection, PIRLS routinely conducts a field test for the purposes of evaluating the measurement properties of the items and practicing the data collection and scoring procedures. In addition to providing important information about how well each assessment item functioned, the PIRLS 2021 field test prompted a number of improvements to the digitalPIRLS player software.

Because the TIMSS & PIRLS International Study Center generally field tests twice the number of passages and items actually required, the PIRLS 2021 field test included a total of 12 newly developed PIRLS passages—6 literary passages and 6 informational texts—with a total of 174 items. In addition, for countries participating in digitalPIRLS, the field test included 2 newly developed ePIRLS online informational tasks with a combined 35 items.

The field test was designed to be conducted in approximately 30 schools in each country and yield at least 200 student responses to each item. The school samples for the PIRLS 2021 Field Test and Data Collection were drawn simultaneously, using the same random sampling procedures to ensure that a school was selected for either the field test or data collection, but not both. For example, if a country needed 150 schools for data collection 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 for the field test from the 180 schools. See Chapter 3 for details about the school and classroom sampling methods and procedures used in PIRLS 2021.

The PIRLS 2021 field tests were conducted in March–June 2020. 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 2021 Field Test

	paperPIRLS	digitalPIRLS
Passages/Tasks	**12**	**14**
Items		
Total	190	230
Responses per item (approx.)	1,000	1,000
Participants		
Countries	8	14
Benchmarking Entities	-	1
Students	5,862	8,381
Teachers	263	458
Schools	145	243

Note: Due to small sample sizes, data for Denmark, Italy, and Lithuania were removed when calculating item statistics for item review.

Exhibit 1.5: PIRLS 2021 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
paperPIRLS						
Literary	6	36	60	96	130	50%
Informational	6	32	62	94	132	50%
Total	**12**	**68**	**122**	**190**	**262**	
digitalPIRLS						
Literary	6	36	60	96	130	41%
Informational	8	48	86	134	188	59%
Total	**14**	**84**	**146**	**230**	**318**	

Exhibit 1.6 PIRLS 2021 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
paperPIRLS					
Focus on and Retrieve Explicitly Stated Information	20	23	43	48	18%
Make Straightforward Inferences	28	34	62	86	33%
Interpret and Integrate Ideas and Information	14	42	56	90	34%
Evaluate and Critique Content and Textual Elements	6	23	29	38	15%
Total	**68**	**122**	**190**	**262**	
digitalPIRLS					
Focus on and Retrieve Explicitly Stated Information	24	26	50	56	18%
Make Straightforward Inferences	38	41	79	109	34%
Interpret and Integrate Ideas and Information	15	51	66	106	33%
Evaluate and Critique Content and Textual Elements	7	28	35	47	15%
Total	**84**	**146**	**230**	**318**	

Due to the ongoing COVID-19 pandemic, not all countries were able to conduct a full-scale field test for PIRLS 2021 as planned. However, most countries unable to participate in the full-scale field test managed to conduct smaller-scale operational field tests at the start of the 2021 school year that served as a "dress-rehearsal" for the data collection and scoring procedures. This was

particularly important for countries participating in PIRLS for the first time, as well as the countries that transitioned to digitalPIRLS.

Developing the Materials for PIRLS 2021 Field Test Scoring Training

In order for field test scoring to occur immediately after the data are collected, 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, seven English-speaking countries participated in a pilot to collect example responses on new constructed response items. Australia, Canada (Ontario), England, Ireland, New Zealand, Singapore, and South Africa administered the newly developed constructed response field test items in a small selection of classrooms with English-speaking students. Approximately 160 sample responses to each new item were collected in September–October 2019.

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	ePIRLS
Passages/Tasks	**12**	**2**
Items		
Total	102	22
Responses per item (approx.)	160	150
Participants		
Countries	Australia, Canada (Ontario), England, Ireland, Singapore	New Zealand, Ireland, Canada, Singapore

In November 2019, staff at the TIMSS & PIRLS International Study Center and a subgroup of three RDG members met to create sets of example and practice responses for 40 PIRLS 2021 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.

The PIRLS 2021 NRCs and their scoring supervisors received scoring training for the field test constructed response items in March 2020 in Belgrade, Serbia as part of the 4th PIRLS 2021

NRC Meeting. The training was conducted by the PIRLS working group consultants and staff at the TIMSS & PIRLS International Study Center. At the scoring training sessions, the trainers explained the purpose of each item and read it aloud. The trainers 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 2021 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.

After the item-by-item review, the TIMSS & PIRLS International Study Center staff collaborated with consultants to assemble the recommended literary and informational texts and item sets, and ePIRLS tasks for the RDG to review. RDG members scrutinized the recommendations for the newly developed assessment materials, reviewing each text and item set as well as scoring guides for content accuracy, clarity, and adherence to the frameworks. In addition, the newly developed texts and items were considered in relation to the trend passages and item sets for overall coherence as a complete assessment.

The 5th NRC meeting held virtually in August 2020 was devoted to reviewing all the recommended passages, tasks, and items for PIRLS 2021. Following this meeting, staff at the TIMSS & PIRLS International Study Center 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 2020.

Exhibit 1.8 includes descriptions of the newly developed PIRLS 2021 passages and tasks, as well as trend passages and tasks from PIRLS 2001, 2006, 2011, and 2016. Exhibit 1.9 presents the word counts and readability scores for the passages.

Exhibit 1.8: Descriptions of the PIRLS 2021 Assessment Texts

	Literary Texts	Informational Texts
Difficult	**Shiny Straw** – This animal story demonstrates heroism and the consequences of a reckless attitude.	**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.
	Ink Drinker* – In this fantasy story, a boy encounters a strange customer in his father's bookstore and as a result has an exciting adventure and learns the joy of reading.	**The World's Bank for Seeds*** – This article describes the world's largest and most secure seed bank located in Svalbard, Norway, and why it is important.
Medium	**The Empty Pot** – This traditional tale set in China has a moral message about the importance of honesty.	**Sharks** – This article presents information about sharks in a variety of formats, using subheadings, a labeled diagram, and photographs.
	Ostrich and The Hat* – In this tale about a father and daughter in Botswana, the father tells his daughter a story about a time he encountered danger and how his hat saved his life.	**Marie Curie Prize-Winning Scientist*** – This biographical text describes Marie Curie's life, her contributions to scientific research, and why she was a role model for other women.
	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.
Easy	**Learning a New Language** – This modern story is about a young girl who struggles to learn to read in the language of her new country.	**The Amazing Octopus** – This article explains the natural habitat and behavior of octopuses, their life in aquariums, and the amazing things they do.
	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.	**Training a Deaf Polar Bear*** – This passage describes how zookeepers worked with a polar bear that was found to be deaf.
	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.

*Text held secure for future assessments.

Exhibit 1.9: Text Word Counts and Readability Scores

Literary	Word Count	Flesch-Kincaid
Shiny Straw	860	5.5
Oliver and the Griffin	896	3.3
Ink Drinker	800	3.7
The Empty Pot	767	4.9
Ostrich and The Hat	788	3.5
Pemba Sherpa	540	2.5
Learning a New Language	561	3.6
The Summer My Father Was Ten	484	4.0
Library Mouse	497	3.1
Informational	**Word Count**	**Flesch-Kincaid**
Where's the Honey?	870	3.2
Icelandic Horses	870	5.0
The World's Bank for Seeds	908	6.8
Sharks	850	7.6
Marie Curie Prize-Winning Scientist	842	8.2
How Did We Learn to Fly?	514	6.3
The Amazing Octopus	674	6.7
Training a Deaf Polar Bear	425	4.0
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 1.10 shows the descriptions of the five ePIRLS tasks assessing online informational reading in PIRLS 2021.

Exhibit 1.10: Descriptions of the PIRLS 2021 ePIRLS Tasks

ePIRLS Tasks in PIRLS 2021
The Legend of Troy – In this history text, students learn about archeology through first investigating the legend of Troy and then learning about the ancient city of Troy.
Zebra and Wildebeest Migration – In this task on animal migration, students focus their learning on Zebras and Wildebeests. As they learn about migration in the Serengeti, students have the opportunity to engage in multiple media forms, including text, charts, interactive images, and animated maps.
Rainforests* – This science text gives students the opportunity to learn about rainforests through text, an interactive map, and a diagram. Students learn about multiple facets of the rainforest, including its structure and the plants and animals that live throughout the different layers.
Oceans – This science-based task gives students the opportunity to learn about the world's oceans. Students are asked to read texts about why oceans are important, including an interview with an oceanographer, as well as use text, videos, and animations to learn about ocean life and habitats and why oceans are threatened.
Voyages of Discovery* – In this task on voyages of discovery, students learn about important sea and space voyages that have changed the way we think about our world. Students have the opportunity to engage in multiple media forms, including animated maps, videos, and pop-ups.

*Task held secure for future assessments.

Distribution of PIRLS 2021 Items by Purpose and Comprehension Process

Exhibits 1.11 and 1.12 present the number of trend and newly developed items as well as the number of score points in PIRLS 2021. 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.

Exhibit 1.11: PIRLS 2021 Achievement Items by Reading Purpose

Reading Purpose	Number of Passages/ Tasks	Trend		New		Total	
		Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points
paperPIRLS							
Literary	9	93 (114)	53%	49 (62)	51%	142 (176)	52%
Informational	9	86 (103)	47%	45 (60)	49%	131 (163)	48%
Total	**18**	**179 (217)**		**94 (122)**		**273 (339)**	
digitalPIRLS							
Literary	9	93 (114)	41%	49 (62)	36%	142 (176)	39%
Informational	14	140 (166)	59%	86 (111)	64%	226 (277)	61%
Total	**23**	**233 (280)**		**135 (173)**		**368 (453)**	

Score points are shown in parentheses.

Exhibit 1.12: PIRLS 2021 Achievement Items by Comprehension Process

Comprehension Process	Trend		New		Total	
	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points	Number of Items	Percentage of Score Points
paperPIRLS						
Focus on & Retrieve Explicitly Stated Information	66 (73)	34%	22 (22)	18%	88 (95)	28%
Make Straightforward Inferences	47 (51)	24%	31 (42)	34%	78 (93)	27%
Interpret & Integrate Ideas and Information	43 (68)	31%	25 (38)	31%	68 (106)	31%
Evaluate & Critique Content and Textual Elements	23 (25)	12%	16 (20)	16%	39 (45)	13%
Total	**179 (217)**		**94 (122)**		**175 (223)**	
digitalPIRLS						
Focus on & Retrieve Explicitly Stated Information	79 (87)	31%	29 (30)	17%	108 (117)	26%
Make Straightforward Inferences	63 (69)	25%	47 (61)	35%	110 (130)	29%
Interpret & Integrate Ideas and Information	57 (88)	31%	37 (54)	31%	94 (142)	31%
Evaluate & Critique Content and Textual Elements	34 (36)	13%	22 (28)	16%	56 (64)	14%
Total	**233 (280)**		**135 (173)**		**368 (453)**	

Score points are shown in parentheses.

Distribution of PIRLS Item Formats within Purpose and Comprehension Process

As described in the *PIRLS 2021 Assessment Frameworks*, 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 1.13 and 1.14 display the number of passages or tasks and items (and score points) by item format for each purpose and comprehension process, respectively.

Exhibit 1.13: PIRLS 2021 Achievement Items by Reading Purpose and Item Format

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		
paperPIRLS								
Literary	9	68 (68)	1 (1)	44 (44)	24 (48)	5 (15)	142 (176)	52%
Informational	9	56 (56)	0 (0)	46 (46)	26 (52)	3 (9)	131 (163)	48%
Total	**18**	**124 (124)**	**1 (1)**	**90 (90)**	**50 (100)**	**8 (24)**	**273 (339)**	
Achieved Percentage of Score Points		**37%**			**63%**			
digitalPIRLS								
Literary	9	68 (68)	1 (1)	44 (44)	24 (48)	5 (15)	142 (176)	39%
Informational	14	98 (98)	4 (7)	79 (79)	42 (84)	3 (9)	226 (277)	61%
Total	**23**	**166 (166)**	**5 (8)**	**123 (123)**	**66 (132)**	**8 (24)**	**368 (453)**	
Achieved Percentage of Score Points		**38%**			**62%**			

Score points are shown in parentheses.

Exhibit 1.14: PIRLS 2021 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		
paperPIRLS							
Focus On and Retrieve Explicitly Stated Information	40 (40)	0 (0)	41 (41)	7 (14)	0 (0)	88 (95)	28%
Make Straightforward Inferences	49 (49)	0 (0)	14 (14)	15 (30)	0 (0)	78 (93)	27%
Interpret and Integrate Ideas and Information	15 (15)	1 (1)	22 (22)	22 (44)	8 (24)	68 (106)	31%
Evaluate and Critique Content and Textual Elements	20 (20)	0 (0)	13 (13)	6 (12)	0 (0)	39 (45)	13%
Total	**124 (124)**	**1 (1)**	**90 (90)**	**50 (100)**	**8 (24)**	**273 (339)**	
Achieved Percentage of Score Points	**37%**			**63%**			
digitalPIRLS							
Focus On and Retrieve Explicitly Stated Information	52 (52)	0 (0)	47 (47)	9 (18)	0 (0)	108 (117)	26%
Make Straightforward Inferences	68 (68)	0 (0)	22 (22)	20 (40)	0 (0)	110 (130)	29%
Interpret and Integrate Ideas and Information	20 (20)	5 (8)	32 (32)	29 (58)	8 (24)	94 (142)	31%
Evaluate and Critique Content and Textual Elements	26 (26)	0 (0)	22 (22)	8 (16)	0 (0)	56 (64)	14%
Total	**166 (166)**	**5 (8)**	**123 (123)**	**66 (132)**	**8 (24)**	**368 (453)**	
Achieved Percentage of Score Points	**38%**			**62%**			

Score points are shown in parentheses.

PIRLS 2021 Constructed Response Scoring Training

In preparation for scoring training for the main data collection, some PIRLS 2021 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 2021 scoring training, the example and practice paper training sets included those used in PIRLS 2016 for the trend items and the updated training sets for the newly developed items selected for PIRLS 2021, resulting in 40 example and practice paper sets for PIRLS and 9 sets for ePIRLS items.

To provide scoring training for all the countries participating in PIRLS 2021, the TIMSS & PIRLS International Study Center conducted two training sessions: 1) Southern Hemisphere countries and their scoring supervisors could attend PIRLS scoring training virtually in November 2020 and 2) Northern Hemisphere countries and their scoring supervisors received scoring training virtually in March 2021 as part of the 6th PIRLS 2016 NRC Meeting.

Exhibit 1.15 shows the number of participants in the two scoring training sessions.

Exhibit 1.15: PIRLS 2021 Scoring Training Participation

Participants	Southern Hemisphere	Northern Hemisphere
Number of Countries	3	54
Number of Benchmarking Entities	0	8
Number of Country Representatives	18	241

Due to the ongoing COVID-19 pandemic, several Southern Hemisphere countries had to postpone their data collection by a year. Singapore, New Zealand, and Australia participated in the Southern Hemisphere scoring training, while Brazil and South Africa participated in the Northern Hemisphere scoring training.

The Process Following Instrument Development

In general, after countries received the international version of the PIRLS 2021 assessment instruments, they began the process of translation and cultural adaptation (some adaptation to local usage typically is necessary even in English-speaking countries), followed by an external review of their translations, and a review of their instrument layout and adaptations (see Chapter 5). At the same time, countries made final arrangements for data collection, including the host of activities necessary to obtain school participation, prepare devices for digitalPIRLS, and implement test administration. However, despite extensive planning and preparation, the circumstances brought on by the COVID-19 pandemic introduced unexpected challenges to PIRLS 2021 Data Collection. In particular, ongoing school closures required some countries to postpone their test administration, which necessitated additional effort in data collection procedures (see Chapter 4).

References

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.

Mullis, I. V. S., & Martin, M. O. (2019). PIRLS 2021 Reading Assessment Framework. In I. V. S. Mullis & M. O. Martin (Eds.), *PIRLS 2021 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/pirls2021/frameworks/