CHAPTER 11

Implementing the PIRLS 2021 Achievement Scaling Methodology

Liqun Yin
Ummugul Bezirhan
Bethany Fishbein
Pierre Foy

Introduction

Continuing the tradition of PIRLS since 2001, PIRLS 2021 is designed to provide international measures of students' reading achievement in their fourth year of schooling and to measure trends in achievement over time. PIRLS is based on a broad definition of reading ability, recognizing different purposes for reading and including a wide variety of texts of varying difficulty levels accessible to students of wide-ranging abilities and cultural contexts (Mullis & Martin, 2019). Given this broad coverage, PIRLS 2021 used a matrix-sampling booklet design such that each student was administered only a subset of the entire PIRLS item pool. PIRLS relied on item response theory (IRT) scaling to provide accurate measures of student proficiency distributions and trends. To provide unbiased estimates of student achievement and its relationship to contextual variables, the PIRLS scaling approach used a latent regression population model with subsequent multiple imputation to obtain plausible values representing proficiency in reading for all students.

This chapter describes the procedures for scaling the PIRLS 2021 achievement data. The Analysis Unit at the TIMSS & PIRLS International Study Center implemented the psychometric scaling and population modeling of the PIRLS 2021 achievement data and conducted related analyses to ensure the quality and validity of the results. A detailed description of the PIRLS 2021 scaling methodology can be found in Chapter 10. Consistent with previous assessments, the 2021 scaling was based on a concurrent calibration of the PIRLS 2021 data jointly with data from the previous PIRLS 2016 cycle for measuring trends. However, the approach for 2021 involved additional psychometric analyses for linking the computer-based data to be reported on the same scale as the paper-based results, including trend measurements from past assessments.

PIRLS 2021 marked the transition from paper-and-pencil to digital format, with about half the countries choosing to administer the digital format (digitalPIRLS) and the other half the paper format (paperPIRLS). To ensure comparability across formats, digitalPIRLS and paperPIRLS shared

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

CHAPTER 11: IMPLEMENTING THE SCALING METHODOLOGY
METHODS AND PROCEDURES: PIRLS 2021 TECHNICAL REPORT
https://doi.org/10.6017/lse.tpisc.tr2103.kb3067   11.1

assessment blocks in terms of reading texts and questions. As described in the PIRLS 2021 Assessment Design (Martin et al., 2019), the PIRLS 2021 group adaptive assessment, both digital and paper versions, consisted of 18 texts and accompanying items, half of them assessing reading for literary experience ("literary") and the other half assessing reading to acquire and use information ("informational"). Countries administering digitalPIRLS additionally administered five ePIRLS tasks that assess informational reading in a simulated internet environment.

Altogether, 57 countries and eight benchmarking entities participated in PIRLS 2021. Participating countries had national samples of at least 150 schools and a minimum of 4,000 students (paperPIRLS) or 5,000 students for digitalPIRLS countries to accommodate integrating ePIRLS. To provide "bridge" data between the paper and digital assessments, digitalPIRLS countries administered paper bridge booklets to an additional sample of 1,500 students, sampling from the same target population as the full digitalPIRLS samples and the same schools to the extent possible. The bridge data form a link between digitalPIRLS countries' computer-based data in 2021 and their paper-based data in 2016, as well as to the paperPIRLS countries in 2021. This bridge allowed for directly comparing the psychometric properties of items in both modes based on equivalent student samples in countries that administered digitalPIRLS. Details on the linking approach are provided in Chapter 10, and selected results on country-level comparisons of the bridge and the digitalPIRLS samples are provided in Chapter 12.

Overview of the PIRLS 2021 Assessments

As described in the PIRLS 2021 Assessment Design (Martin et al., 2019), PIRLS 2021 adopted a unified group adaptive assessment design to address the need for a broader range of assessment difficulty and better targeting of student ability to improve measurement. PIRLS previously introduced less difficult versions of the reading assessment, starting with prePIRLS in 2011 and following up with the PIRLS Literacy assessment in 2016. While these parallel assessments successfully expanded the PIRLS coverage at the lower end of the ability distribution, they did not address the need for coverage at the higher end.

The PIRLS 2021 group adaptive assessment consisted of 18 assessment blocks in the form of text and item sets. Half of the 18 texts assessed the literary reading purpose and the other half assessed informational reading. Based on average item difficulties, the 18 texts were grouped into three levels of difficulty—easy, medium, and difficult—with three literary and three informational texts at each difficulty level. The 18 texts were assembled into 18 booklets, with each booklet containing one literary text and one informational text. Each text appeared in two booklets and was paired with a different text each time. The booklets were divided into two levels of difficulty: nine more difficult booklets composed of either two difficult texts or one medium and one difficult text, and nine less difficult booklets composed of two easy texts or one easy and one medium

text. All 18 booklets were distributed in all countries, but with varying proportions of the more and less difficult booklets depending on the average reading ability of the national student populations, which was determined based on performance in prior PIRLS assessments. Chapter 9 reports the proportion of more and less difficult booklets administered in each PIRLS 2021 country.

Of the 18 text and item sets in the group adaptive design, eight also were administered in 2016, including two texts shared between PIRLS 2016 and PIRLS Literacy 2016, providing the basis for trend measurement. Four text and item sets also were administered in PIRLS Literacy 2016. The other six were developed and field tested for first-time use in PIRLS 2021.

The eight trend texts and their items were developed when paper-and-pencil was the sole mode of administration. For PIRLS 2021, a digital version was developed for each text, retaining the look and feel of the paper versions as much as possible. Four assessment blocks from PIRLS Literacy 2016 were also adapted for digitalPIRLS, but underwent additional format changes. In PIRLS Literacy assessment blocks on paper, the text and items were split up into portions such that each item corresponded to only a small portion of the full text shown on the opposite page of the open booklet. When the PIRLS Literacy 2016 paper texts and items were converted to digital versions, the presentation had to be changed to harmonize the functionality of PIRLS Literacy and "regular" PIRLS passages for inclusion in the digitalPIRLS assessment. This was required to present test takers with a uniform interface when answering items on the computer, without presenting items directly next to the portion of text to which they pertain. For this reason, and also because they were only administered to a limited number of countries in PIRLS Literacy 2016, these four passages and their items were not considered part of the trend materials.

PIRLS 2021 integrated five ePIRLS informational tasks with digitalPIRLS to represent a more comprehensive picture of informational reading as PIRLS transitioned to a computer-based assessment. ePIRLS was first introduced in 2016 as an extension of the PIRLS reading framework with online informational reading. The ePIRLS tasks simulate an authentic online reading environment in which students gather information, using links and tabs to navigate through texts and graphics to accomplish school-based projects or reports, with an emphasis placed on assessing reading comprehension. Among the five ePIRLS tasks, three were already administered in 2016. Two tasks were developed for first-time use in PIRLS 2021. Because online reading tends to be multimodal in the ways they present information and contain practical features (such as video clips and animated graphics) that are not possible to reproduce in print format, ePIRLS was available only in digital format. For digitalPIRLS countries, a subset of students was presented with ePIRLS material, either two ePIRLS tasks or one digitalPIRLS informational passage followed by one ePIRLS task, as described in the PIRLS 2021 Assessment Design (Martin et al., 2019).

The paperPIRLS Assessment

Countries participating in paperPIRLS administered the 18 booklets comprised of 18 text and item sets under the group adaptive assessment design described above. Of the 57 countries and eight benchmarking entities participating in PIRLS 2021, 31 countries and one benchmarking entity administered the paper assessment.

The digitalPIRLS Assessment

The remaining 26 countries and seven benchmarking entities administered the digitalPIRLS assessment. The digitalPIRLS assessment consisted of three types of digital booklets that included the PIRLS texts and the ePIRLS tasks. The regular digitalPIRLS booklet design followed the group adaptive design with the 18 texts assembled into 18 "digitalPIRLS" booklets (Booklets 1 to 18). The five ePIRLS tasks were assembled into 20 distinct "ePIRLS" booklets (Booklets 19 to 38), the results of all possible pairings of the five ePIRLS tasks. Finally, each of the nine PIRLS informational texts was paired with each of the five ePIRLS tasks, resulting in 45 "hybrid" booklets (Booklets 39 to 83), with one digitalPIRLS informational text followed by one ePIRLS task.

The regular digitalPIRLS booklets, ePIRLS booklets, and hybrid booklets were included in the booklet assignment rotation scheme in digitalPIRLS countries. To avoid the imbalance of the number of regular digitalPIRLS booklets assigned relative to the number in paperPIRLS countries, digitalPIRLS used a 27-booklet rotation that included all 18 digitalPIRLS booklets, six of the ePIRLS booklets on a rotating basis, and three of the hybrid booklets, also on a rotating basis. As a result, two-thirds of students received a regular booklet, two-ninths received an ePIRLS booklet, and one-ninth received a hybrid booklet (Martin et al., 2019).

The digitalPIRLS Bridge Booklets

All 26 countries and one benchmarking entity participating in digitalPIRLS also administered a set of eight paper bridge booklets consisting entirely of the paper trend texts. Seven of these booklets were also administered in 2016, and one contained two texts also administered in 2016 but in a different combination. The data from these paper bridge booklets were used to link the digitalPIRLS assessment to the paperPIRLS assessment and to the PIRLS trend scale, relying on equivalent populations between the digitalPIRLS and bridge samples. The data also provided countries with information on the behavior of mode format differences in their countries (see Chapter 12).

The booklets containing paper bridge texts and items were administered to national samples of 1,500 students selected through an equivalent groups design, ensuring their randomness and equivalence with the national digitalPIRLS samples of their respective countries. The equivalent groups design involved randomly assigning classes, often in the same schools, to either the digital or paper-based assessment format. While a sizeable overlap within schools is desirable to strengthen the random equivalence of the bridge and digital samples, all national bridge and

digital samples were sampled from the same national target populations and used the same sample designs to establish their random equivalence.

The Linking Design between Assessments

The linking design between the paper and digital versions of PIRLS 2021 is shown in Exhibit 11.1, illustrating which reading texts and tasks were present in each set of assessment booklets. Exhibit 11.2 reports the number of items in each assessment.

Exhibit 11.1: Texts in the PIRLS 2021 Assessments

Reading Purpose	Difficulty Level	Text	paperPIRLS Assessment Booklets	digitalPIRLS Bridge Booklets	digitalPIRLS Assessment		
					digitalPIRLS Booklets	Hybrid Booklets	ePIRLS Booklets
Literary	Difficult	Shiny Straw (06)	●	●	●		
		Oliver and the Griffin (16)	●	●	●		
		The Ink Drinker (21)	●	●			
	Medium	The Empty Pot (11)	●	●	●		
		Pemba Sherpa (16)	●	●	●		
		Ostrich and the Hat (21)	●	●			
	Easy	The Summer My Father Was 10 (11)	●	●			
		Library Mouse (16)	●	●			
		Learning a New Language (21)	●	●			
Informational	Difficult	Where's the Honey? (11)	●	●	●	●	
		Icelandic Horses (16)	●	●	●	●	
		The World's Bank for Seeds (21)	●	●	●		
	Medium	Sharks (06)	●	●	●	●	
		How Did We Learn to Fly? (16)	●	●	●	●	
		Marie Curie Prize-Winning Scientist (21)	●	●	●		
	Easy	Training a Deaf Polar Bear (11)	●	●	●		
		Hungry Plant (16)	●	●	●		
		The Amazing Octopus (21)	●	●	●		

Reading Purpose	Difficulty Level	Text	paperPIRLS Assessment Booklets	digitalPIRLS Bridge Booklets	digitalPIRLS Assessment		
					digitalPIRLS Booklets	Hybrid Booklets	ePIRLS Booklets
Online Informational		Rainforests (16)				●	●
		The Legend of Troy (16)				●	●
		Zebra and Wildebeest Migration (16)				●	●
		Oceans (21)				●	●
		Voyages of Discovery (21)				●	●

( ) Number in parentheses indicates the assessment year in which the passage was first introduced.

Exhibit 11.2: Number of Items in the PIRLS 2021 Assessments by Reading Purpose

Assessment		Literary	Informational	Total
paperPIRLS		141	131	272
Bridge		61	56	117
digitalPIRLS	digitalPIRLS Items	141	131	272
	ePIRLS Items	0	91	91

Countries Participating in PIRLS 2021

Exhibit 11.3 shows the number of participating countries and benchmarking participants in PIRLS 2021 for paperPIRLS and digitalPIRLS. In PIRLS 2021, 31 countries and one benchmarking entity administered the full assessment using paper booklets, and 26 countries and seven benchmarking entities transitioned to digital assessment. All digitalPIRLS participants also administered the bridge booklets.

Because of delays in test administration, only trend countries that participated in the previous PIRLS 2016 cycle and administered the PIRLS 2021 assessment to the fourth grade student cohort at the end of the school year in 2020 or 2021 were included in the concurrent item calibration model. The COVID-19 pandemic introduced many challenges to the PIRLS 2021 data collection, requiring a more targeted group of countries contribute to setting the scale. Most countries managed to administer the PIRLS 2021 assessment to the targeted fourth grade students at the end of the school year in 2020 or 2021, according to the original plan. However, 14 countries and three benchmarking participants delayed assessment of the cohort of fourth grade students until the beginning of the next school year in the fifth grade. These countries had samples of markedly older students compared to their PIRLS 2016 counterparts. Six countries and one benchmarking participant assessed the next fourth grade cohort of students one year later. Three of these six

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

CHAPTER 11: IMPLEMENTING THE SCALING METHODOLOGY
METHODS AND PROCEDURES: PIRLS 2021 TECHNICAL REPORT   11.6

countries did not collect data until 2022. More on how the delays affected samples and reporting can be found in Chapter 8 and Chapter 12.

Altogether, 11 paperPIRLS trend countries and 16 digitalPIRLS trend countries were included in the paper concurrent calibration model (described later in this chapter). The 16 digitalPIRLS trend countries also were included in the digital item calibration model for the population-based linking and ePIRLS item calibration.

Exhibit 11.3: **Number of Countries Participating in PIRLS 2021 by Data Collection Period**

	All Countries	Trend Countries (2016)	Calibration Countries*	Benchmarking Participants
paperPIRLS	31	19	11	1
digitalPIRLS	26	24	16	7
Total	**57**	**43**	**27**	**8**

* Because of delays in test administration, only trend countries that participated in the previous PIRLS 2016 cycle and administered the PIRLS 2021 assessment to the fourth grade student cohort at the end of the school year in 2020 or 2021 were included in the item calibration models and the calculation of scale transformation constants.

Overview of Scaling the PIRLS 2021 Achievement Data

The PIRLS reporting metric was originally established in PIRLS 2001 by setting the mean of the national average scores for all countries that participated in 2001 to 500 and the standard deviation to 100. Successive framework updates and item release policies changed the composition of subsequent PIRLS rounds in 2006, 2011, 2016 and 2021. To enable measurement of trends over time, achievement data from successive PIRLS assessments were transformed to that same metric by linking each new data set to the immediate predecessor's scale. This was done by concurrently scaling the data from each successive assessment, here 2021, with the data from the previous assessment, here 2016—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 means that PIRLS 2021 is linked to 2001 through several steps that involve changes over time in item selection and domain coverage due to framework updates. This procedure enables PIRLS to measure trends across all five assessment cycles: 2021 linked to 2016, 2016 linked to 2011, 2011 to 2006, and 2006 to 2001.

Consequently, item parameters were estimated for all paper items through a concurrent calibration of the data from the 2021 and 2016 assessments. The digitalPIRLS data were linked with the paper data with the data collected from the bridge samples, under a commonly used linking approach referred to as a "randomly equivalent samples" design (Dorans & Puhan, 2017; Haberman, 2015; Kolen & Brennan, 2014; M. von Davier & A. von Davier, 2007). To estimate

the digital item parameters, this population-based equivalent groups linking approach involved estimating all item parameters freely without assuming any item equivalence between the two modes of administration, while setting the scale through equivalency assumptions regarding the population distributions of digital- and paper-based samples.

To avoid any potential influence of the ePIRLS data on the joint paperPIRLS and digitalPIRLS scale, the ePIRLS items were linked to the same scale in a separate calibration step utilizing fixed item parameter linking (e.g., Kim, 2006; Vale, 1986; M von Davier & A. von Davier, 2007). Alongside the digitalPIRLS data, the ePIRLS item parameters were freely estimated with the digitalPIRLS item parameters fixed to equal the values estimated in the previous step. Linking ePIRLS relied on sample equivalence and the presence of common items among the regular digital booklets, hybrid booklets, and ePIRLS booklets (see Exhibit 11.1).

After item calibration, the data from paperPIRLS, digitalPIRLS, and ePIRLS underwent further analysis steps using latent regression models (e.g. Mislevy, 1984; Mislevy & Sheehan, 1987; von Davier et al., 2006; von Davier et al., 2009) to generate plausible values (imputations) for overall reading achievement, as well as for the reading purposes and the comprehension processes. The use of a latent regression IRT model to generate plausible values is sometimes called *conditioning* and uses contextual data from the student and parent questionnaires as predictors of achievement to increase the reliability of the imputations. Finally, the plausible values were placed on the PIRLS trend reporting metric through a series of linear transformations.

Three main objectives were addressed in scaling the PIRLS 2021 achievement data: 1) maintain trends with past PIRLS assessments through concurrent calibration with the 2016 data, 2) report paper and digital results on the same PIRLS trend scale, and 3) integrate ePIRLS data with digitalPIRLS data on the PIRLS trend scale. To achieve these goals, the TIMSS & PIRLS International Study Center conducted four major analysis phases. To minimize sources of interaction across data sources, a stepwise approach was adopted to scale all the PIRLS 2021 data within these four phases—first scaling paperPIRLS data, then digitalPIRLS, and finally ePIRLS.

1. **Item calibration:** In the first phase, item parameters were estimated using multiple-group IRT models for all PIRLS 2021 items in three separate steps for paper items, digital items, and ePIRLS items, respectively.

2. **Principal component analysis:** In the second phase, principal components were extracted from context data for each country for use in population modeling.

3. **Latent regression population modeling:** In the third phase, latent regression models were estimated (conditioning) for each country's data to draw plausible values (PVs) of overall reading achievement, as well as for reading purposes and comprehension processes.

4. **Scale transformation:** Finally, the generated plausible values were placed on the PIRLS metric using linear transformations to report trends from previous assessments and to link the digital data with the paper data.

The PIRLS 2021 scaling procedures are described in this chapter under four sections according to these phases. Several quality checks and analyses were conducted iteratively throughout the scaling process. These analyses and their outcomes are described later in the "Validating the PIRLS 2021 Achievement Results" section of this chapter. IRT models, population models, and other theoretical foundations for these procedures are described in Chapter 10.

Before scaling the PIRLS 2021 achievement data, PIRLS conducted an extensive item-by-item review of classical item statistics for all countries to evaluate the quality of the assessment items and to identify any unexpected or problematic item behaviors. This review also included analyses of change with respect to percent correct and partial credit percentages, omit rates, item discrimination, and other classical item statistics for trend items relative to the 2016 assessment. These item review activities are described in Chapter 9.

Treatment of Item-Level Non-Response (Omitted and Not-Reached)

Given the matrix-sampling design used by PIRLS, whereby a student is administered only a subset of the 18 PIRLS texts and five ePIRLS tasks, most item responses are missing by design for any given student. Students were assigned booklets randomly according to the design described in the PIRLS framework, so that the data due to the matrix-sampling are missing at random (Little & Rubin, 1987; Rubin, 1976) and ignorable in the analysis. 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, does not have sufficient time to attempt the item, or other reasons. PIRLS considers an item to be "not reached" when—within the first or second part of a booklet—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. All other skipped responses would be considered "omitted."

The TIMSS & PIRLS International Study Center introduced a mechanism for the treatment of item non-response for scaling the PIRLS 2021 achievement data that corrects for omitted or not-reached responses based on the statistical dependency of missingness on achievement. This approach is based on the strength of the evidence of missing data for estimating achievement. It assumes that non-response occurs non-randomly, but does not assume it occurs exclusively due to low ability of the students. This strength-of-evidence approach avoids bias due to treating non-response deterministically as if they were incorrect responses (Glas & Pimentel, 2006; Moustaki & Knott, 2000; Rose et al., 2010, 2017).

In the PIRLS 2021 achievement scaling, both omitted and not-reached responses were ignored, or treated as "not administered" for estimating item parameters. For generating student plausible values, to account for the non-randomly missing item responses, a set of *non-response indicators* were created and their IRT parameters were estimated similar to items. The variables indicated whether each student answered all items (1) or had at least one missing response (0).

Two sets of four non-response indicators were created for each student defined by item format (multiple-choice and constructed-response) and PIRLS reading subdomain. The first set of four *purpose non-response indicators* were for the two reading purposes (Literary and Informational) for multiple-choice (MC) and constructed-response (CR) items, respectively. Students taking hybrid and ePIRLS booklets only answered items for the Informational purpose. Therefore, the four "purpose" non-response indicators instead were created according to the position of the text or task in the booklet (part 1 and part 2). The second set of four *process non-response indicators* were for the two comprehension processes (Retrieving and Straightforward Inferencing and Interpreting, Integrating, and Evaluating).

To mitigate any potential effects of the non-response indicators on item parameter estimation, a three-step approach was adopted for estimating achievement item parameters and non-response indicators in all calibration steps. First, only achievement items were included in the calibration for item parameter estimation. In a second step, a three-parameter (3PL) IRT model was applied to estimate the parameters for the four purpose non-response indicators, with all achievement item parameters fixed to equal the values estimated in the first step. Third, a 3PL model was applied to estimate the parameters for the four process indicators with all achievement item parameters fixed.

The non-response indicators were then included alongside achievement items as predictors in the population models for generating student plausible values. The purpose non-response indicators were used for generating plausible values in overall reading as well as for the two reading purposes. The process indicators were used for generating plausible values for the two comprehension processes. This approach incorporates non-response into the generation of plausible values according to the dependency of missingness on achievement, or to the extent that missingness and achievement are (negatively) correlated in the population.

Phase 1: Item Calibration

Item calibration was conducted at the TIMSS & PIRLS International Study Center using an open-source package MIRT (Chalmers, 2012), available in the R statistical programming language (R Core Team, 2020). To achieve the goals outlined in the overview, three calibration models were employed through a stepwise approach to estimate all PIRLS 2021 item parameters:

1. **Paper concurrent calibration:** Relying on the usual PIRLS concurrent calibration approach (e.g., Foy & Yin, 2017), extended for multiple populations, data from the 11 paperPIRLS calibration countries and bridge data from the 16 digitalPIRLS calibration countries (see Exhibit 11.3) were scaled along with their PIRLS 2016 data to estimate item parameters for all the items presented in the paper format.

2. **Digital item calibration:** Relying on the randomly equivalent bridge and digital samples drawn from the same target populations, this calibration and linking model combined the data from the regular digitalPIRLS booklets of the 16 digitalPIRLS calibration countries, with all available paper data from the previous paper concurrent calibration model. The multiple-group model estimated all paper and digital item parameters freely, treating the PIRLS 2021 digital items as different from paper items.

3. **ePIRLS item calibration:** This item calibration put the ePIRLS items on the same scale as the digitalPIRLS data. It utilized the full digital samples from the 16 digital trend calibration countries and fixed the digitalPIRLS item parameters estimated from the digital calibration to estimate ePIRLS item parameters. Linking ePIRLS to digitalPIRLS relied on sample equivalence arising from booklets spiraled within classes and the presence of common items among the three types of booklets in the digitalPIRLS administration.

These calibration models are described in detail in the following sections. After each item calibration step, extensive checks were performed to evaluate the fit of the estimated item parameters to the response data. More details about model fit are provided in the "Validating the PIRLS 2021 Achievement Results" section later in this chapter.

Paper Concurrent Calibration

Item parameters were estimated for all paper items through a concurrent calibration of the data from the 2021 and 2016 assessments. In linking these two successive assessments, concurrent calibration relies on having a large proportion of trend items—items that are retained from one assessment to the next. 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 on a common scale using the item parameters from the concurrent calibration. The difference between the two assessment distributions is the trend measure between the previous and current assessments, although not yet on the PIRLS trend reporting metric until a set of transformations is applied.

Exhibit 11.4 illustrates the general structure of the PIRLS 2021 concurrent calibration model to estimate the paper item parameters. The upper panel of the exhibit, labeled "PIRLS 2016

Calibration," represents the PIRLS 2016 data from the PIRLS 2021 trend countries and the overall student ability distribution, shown on the right, which was estimated for this population in PIRLS 2016. The lower panel of Exhibit 11.4 is labeled "PIRLS 2021 Concurrent Calibration" and illustrates the full array of paperPIRLS and bridge data included in the PIRLS 2021 concurrent calibration model. This included the PIRLS 2016 data from all the PIRLS 2021 calibration trend countries, as well as all 2021 data from paperPIRLS calibration countries and bridge data from digitalPIRLS calibration countries.

Exhibit 11.4: **PIRLS 2021 Paper Concurrent Calibration Model**
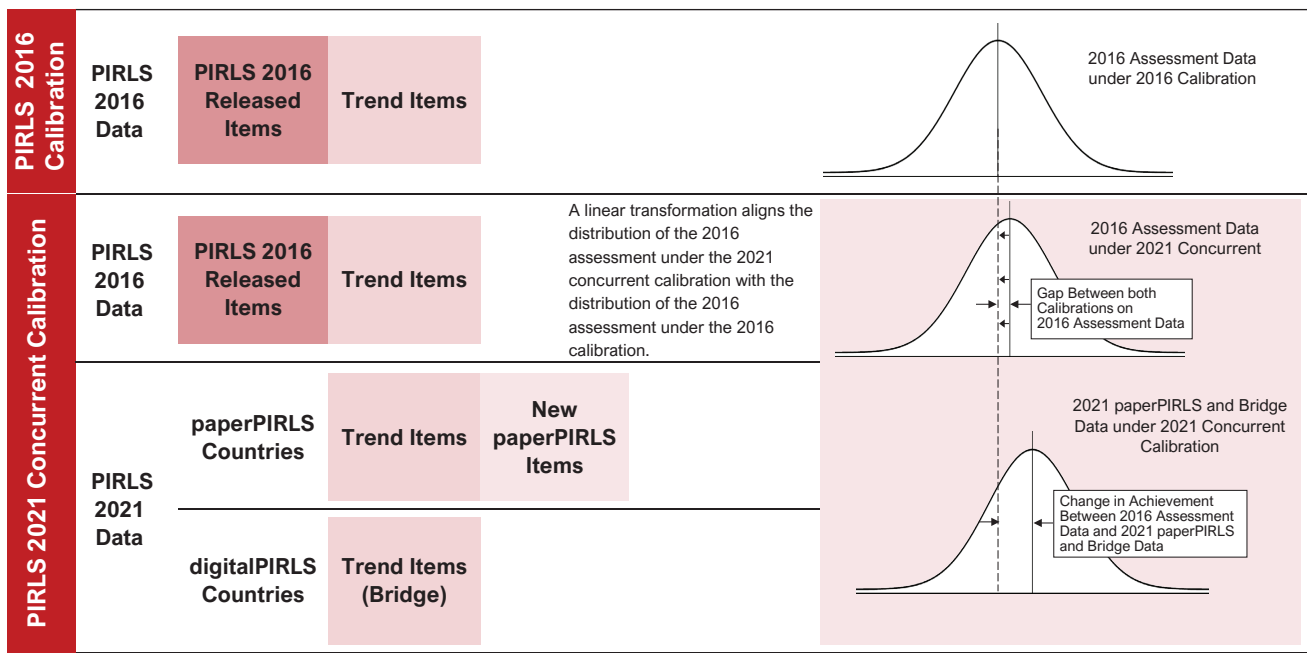
Exhibit 11.5 lists the number of paper items included in the concurrent calibration by item type and by reading purposes and comprehension processes.

Exhibit 11.5: Items for the PIRLS 2021 Paper Concurrent Calibration

Item Type	Points	Items Released in 2016		Items Common in 2016 & 2021		Items Introduced in 2021		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	28	28	57	57	66	66	151	151
Constructed Response	1	16	16	33	33	58	58	107	107
	2	10	20	22	44	28	56	60	120
	3	3	9	5	15	3	9	11	33
Total		**57**	**73**	**117**	**149**	**155**	**189**	**329**	**411**

Items by Reading Purpose and Comprehension Process

Reading Purpose	Items Released in 2016		Items Common in 2016 & 2021		Items Introduced in 2021		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Literary Experience	29	35	61	78	80	97	170	210
Acquire and Use Information	28	38	56	71	75	92	159	201

Comprehension Process	Items Introduced in 2016		Items Common in 2016 & 2021		Items Introduced in 2021		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Retrieving and Straightforward Inferencing	34	39	68	76	97	111	199	226
Interpreting, Integrating, and Evaluating	23	34	49	73	58	78	130	185
Total	**57**	**73**	**117**	**149**	**155**	**189**	**329**	**411**

Exhibit 11.6 shows the sample sizes for scaling the paperPIRLS data, for both item calibration and proficiency estimation. All student samples were weighted so that each country contributed equally to the item calibration. The 2021 concurrent calibration model included data only from trend calibration countries (see Exhibit 11.3). As a result, 27 countries, 11 paperPIRLS trend countries and 16 digitalPIRLS trend countries, contributed to the concurrent calibration. In total, the paper concurrent calibration included 141,645 students from the 2016 assessment and 79,861 students from the 2021 assessment, including bridge data from the digitalPIRLS calibration countries. Student proficiency was estimated for a total of 57 countries and two benchmarking participants for the paperPIRLS and bridge data.

Exhibit 11.6: Sample Sizes for the PIRLS 2021 Paper Concurrent Calibration

Country	Item Calibration			Proficiency Estimation		
	Paper 2016	Paper 2021	Bridge 2021	Paper 2016	Paper 2021	Bridge 2021
Albania	—	—	—	—	4,213	—
Australia	6,341	5,487	—	6,341	5,487	—
Austria	4,360	4,806	—	4,360	4,806	—
Azerbaijan	5,994	5,209	—	5,994	5,209	—
Bahrain	—	—	—	5,480	5,208	—
Belgium (Flemish)	5,198	—	1,623	5,198	—	1,623
Belgium (French)	4,623	4,279	—	4,623	4,279	—
Brazil	—	—	—	—	4,941	—
Bulgaria	4,281	4,043	—	4,281	4,043	—
Chinese Taipei	4,326	—	1,669	4,326	—	1,669
Croatia	—	—	—	—	—	1,226
Cyprus	—	—	—	—	4,589	—
Czech Republic	5,537	—	1,906	5,537	—	1,906
Denmark	3,508	—	1,403	3,508	—	1,403
Egypt	—	—	—	—	7,979	—
England	—	—	—	5,095	4,150	—
Finland	4,896	—	2,069	4,896	—	2,069
France	4,767	5,339	—	4,767	5,339	—
Georgia	—	—	—	5,741	5,241	—
Germany	3,959	—	1,343	3,959	—	1,343
Hong Kong SAR	3,349	3,830	—	3,349	3,830	—
Hungary	—	—	—	4,623	—	1,697
Iran, Islamic Rep. of	—	—	—	4,385	5,962	—
Ireland	—	—	—	4,607	4,663	—
Israel	—	—	—	4,041	—	1,780
Italy	3,940	—	1,979	3,940	—	1,979
Jordan	—	—	—	—	6,150	—
Kazakhstan	—	—	—	4,925	—	3,207
Kosovo	—	—	—	—	4,557	—
Latvia	—	—	—	4,157	4,369	—
Lithuania	—	—	—	4,317	—	1,519
Macao SAR	4,059	5,093	—	4,059	5,093	—

Country	Item Calibration			Proficiency Estimation		
	Paper 2016	Paper 2021	Bridge 2021	Paper 2016	Paper 2021	Bridge 2021
Malta	—	—	—	—	—	835
Montenegro	—	—	—	—	4,489	—
Morocco	—	—	—	5,489	7,017	—
Netherlands	4,206	3,627	—	4,206	4,313	—
New Zealand	5,645	—	2,221	5,645	—	2,221
North Macedonia	—	—	—	—	2,929	—
Northern Ireland	—	—	—	3,693	4,050	—
Norway (5)	4,232	—	1,673	4,232	—	1,673
Oman	9,234	5,321	—	9,234	5,321	—
Poland	4,413	4,179	—	4,413	4,179	—
Portugal	4,642	—	2,098	4,642	—	2,098
Qatar	—	—	—	9,077	—	1,343
Russian Federation	4,577	—	2,187	4,577	—	2,187
Saudi Arabia	—	—	—	4,741	—	1,872
Serbia	—	—	—	—	4,037	—
Singapore	6,488	—	1,988	6,488	—	1,988
Slovak Republic	5,451	—	1,640	5,451	—	1,640
Slovenia	4,499	—	1,414	4,499	—	1,414
South Africa	—	—	—	—	12,422	—
Spain	14,595	—	1,572	14,595	—	1,572
Sweden	4,525	—	1,863	4,525	—	1,863
Turkiye	—	—	—	—	6,032	—
United Arab Emirates	—	—	—	16,471	—	1,990
United States	—	—	—	4,425	—	1,657
Uzbekistan	—	—	—	—	5,846	—
Benchmarking Participants						
Moscow City, Russian Fed.	—	—	—	4,289	—	1,695
South Africa (6)	—	—	—	—	9,317	—
TOTAL	**141,645**	**51,213**	**28,648**	**237,201**	**170,060**	**47,469**

A multiple-group IRT model was utilized for item calibration, specifying country-by-cycle groups and resulting in 54 groups for the concurrent calibration. While the item parameters remain equal across groups, the model estimated distinct ability distributions by country and cycle to properly account for achievement differences. The item parameters estimated from the concurrent calibration for all paper items are presented in Appendix 11A.

The item parameters estimated from this concurrent calibration were then used to estimate student proficiency for all countries and benchmarking entities participating in the paperPIRLS 2021 and bridge assessments. These item parameters also were used to estimate student proficiency for PIRLS 2016 in order to establish the scale transformation necessary to place the PIRLS 2021 results on the PIRLS trend scale.

Digital Item Calibration

The objective of this calibration step was to estimate digitalPIRLS item parameters, linked to the paperPIRLS assessment, and unaffected by any potential influences of the ePIRLS data on the digital item parameter estimation. To achieve this, the randomly equivalent samples linking approach, described in Chapter 10, was employed to link the paper and digital assessments. The linking approach assumes sample equivalence, and evidence supporting this equivalency is presented in Chapter 12.

Exhibit 11.7 shows the number of items included in the digital calibration model. Only the regular digitalPIRLS items, which had counterparts in the paperPIRLS assessment, were included in this calibration model. The data from digitalPIRLS Booklets 1–18 of the 16 trend calibration countries were combined with data from the paper concurrent calibration to estimate parameters for all 601 items without any further constraints, treating the digital items as different from their paper counterparts.

Exhibit 11.7: **Items for the PIRLS 2021 Digital Calibration**

Item Type	Points	Paper Items						Digital Items	
		Items Released in 2016		Items Common in 2016 & 2021		Items Introduced in 2021			
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple Choice	1	28	28	57	57	66	66	123	123
Constructed Response	1	16	16	33	33	58	58	91	91
	2	10	20	22	44	28	56	50	100
	3	3	9	5	15	3	9	8	24
Total		**57**	**73**	**117**	**149**	**155**	**189**	**272**	**338**

Exhibit 11.8 shows the sample sizes for the digital item calibration. The digital calibration used the same data as the paper concurrent calibration (Exhibit 11.6), plus the data from the regular digitalPIRLS booklets from the digitalPIRLS trend calibration countries. Along with the 221,506 students from the paper concurrent calibration, the digital calibration also included 61,256 students from the 16 digitalPIRLS trend calibration countries who were administered one of the 18 digitalPIRLS booklets.

Exhibit 11.8: Sample Sizes for the PIRLS 2021 Digital Calibration

Country	Item Calibration				Proficiency Estimation			
	Paper 2016	Paper 2021	Bridge 2021	Digital 2021	Paper 2016	Paper 2021	Bridge 2021	Digital 2021
Albania	—	—	—	—	—	4,213	—	—
Australia	6,341	5,487	—	—	6,341	5,487	—	—
Austria	4,360	4,806	—	—	4,360	4,806	—	—
Azerbaijan	5,994	5,209	—	—	5,994	5,209	—	—
Bahrain	—	—	—	—	5,480	5,208	—	—
Belgium (Flemish)	5,198	—	1,623	3,408	5,198	—	1,623	3,408
Belgium (French)	4,623	4,279	—	—	4,623	4,279	—	—
Brazil	—	—	—	—	—	4,941	—	—
Bulgaria	4,281	4,043	—	—	4,281	4,043	—	—
Chinese Taipei	4,326	—	1,669	3,696	4,326	—	1,669	3,696
Croatia	—	—	—	—	—	—	1,226	2,623
Cyprus	—	—	—	—	—	4,589	—	—
Czech Republic	5,537	—	1,906	4,388	5,537	—	1,906	4,388
Denmark	3,508	—	1,403	3,220	3,508	—	1,403	3,220
Egypt	—	—	—	—	—	7,979	—	—
England	—	—	—	—	5,095	4,150	—	—
Finland	4,896	—	2,069	4,676	4,896	—	2,069	4,676
France	4,767	5,339	—	—	4,767	5,339	—	—
Georgia	—	—	—	—	5,741	5,241	—	—
Germany	3,959	—	1,343	3,068	3,959	—	1,343	3,068
Hong Kong SAR	3,349	3,830	—	—	3,349	3,830	—	—
Hungary	—	—	—	—	4,623	—	1,697	3,545
Iran, Islamic Rep. of	—	—	—	—	4,385	5,962	—	—
Ireland	—	—	—	—	4,607	4,663	—	—
Israel	—	—	—	—	4,041	—	1,780	3,267

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

Country	Item Calibration				Proficiency Estimation			
	Paper 2016	Paper 2021	Bridge 2021	Digital 2021	Paper 2016	Paper 2021	Bridge 2021	Digital 2021
Italy	3,940	—	1,979	3,628	3,940	—	1,979	3,628
Jordan	—	—	—	—	—	6,150	—	—
Kazakhstan	—	—	—	—	4,925	—	3,207	4,668
Kosovo	—	—	—	—	—	4,557	—	—
Latvia	—	—	—	—	4,157	4,369	—	—
Lithuania	—	—	—	—	4,317	—	1,519	3,097
Macao SAR	4,059	5,093	—	—	4,059	5,093	—	—
Malta	—	—	—	—	—	—	835	2,050
Montenegro	—	—	—	—	—	4,489	—	—
Morocco	—	—	—	—	5,489	7,017	—	—
Netherlands	4,206	3,627	—	—	4,206	4,313	—	—
New Zealand	5,645	—	2,221	3,707	5,645	—	2,221	3,707
North Macedonia	—	—	—	—	—	2,929	—	—
Northern Ireland	—	—	—	—	3,693	4,050	—	—
Norway (5)	4,232	—	1,673	3,594	4,232	—	1,673	3,594
Oman	9,234	5,321	—	—	9,234	5,321	—	—
Poland	4,413	4,179	—	—	4,413	4,179	—	—
Portugal	4,642	—	2,098	4,117	4,642	—	2,098	4,117
Qatar	—	—	—	—	9,077	—	1,343	3,497
Russian Federation	4,577	—	2,187	3,488	4,577	—	2,187	3,488
Saudi Arabia	—	—	—	—	4,741	—	1,872	3,186
Serbia	—	—	—	—	—	4,037	—	—
Singapore	6,488	—	1,988	4,482	6,488	—	1,988	4,482
Slovak Republic	5,451	—	1,640	3,220	5,451	—	1,640	3,220
Slovenia	4,499	—	1,414	3,416	4,499	—	1,414	3,416
South Africa	—	—	—	—	—	12,422	—	—
Spain	14,595	—	1,572	5,691	14,595	—	1,572	5,691
Sweden	4,525	—	1,863	3,457	4,525	—	1,863	3,457
Turkiye	—	—	—	—	—	6,032	—	—
United Arab Emirates	—	—	—	—	16,471	—	1,990	18,322
United States	—	—	—	—	4,425	—	1,657	—
Uzbekistan	—	—	—	—	—	5,846	—	—

Country	Item Calibration				Proficiency Estimation			
	Paper 2016	Paper 2021	Bridge 2021	Digital 2021	Paper 2016	Paper 2021	Bridge 2021	Digital 2021
Benchmarking Participants								
Alberta, Canada	–	–	–	–	–	–	–	2,006
British Columbia, Canada	–	–	–	–	–	–	–	3,111
Newfoundland & Labrador, Can.	–	–	–	–	–	–	–	1,623
Quebec, Canada	–	–	–	–	–	–	–	2,496
Moscow City, Russian Fed.	–	–	–	–	4,289	–	1,695	3,837
South Africa (6)	–	–	–	–	–	9,317	–	–
Abu Dhabi, UAE	–	–	–	–	–	–	–	6,941
Dubai, UAE	–	–	–	–	–	–	–	5,137
TOTAL	**141,645**	**51,213**	**28,648**	**61,256**	**237,201**	**170,060**	**47,469**	**130,662**

Similar to the concurrent calibration, country-by-cycle groups were defined for the multiple-group IRT model for the digital data. Because each country's respondents from the digitalPIRLS and bridge samples were drawn from the same population and treated as equivalent samples, bridge and digital samples were assumed to be one group. This resulted in 54 groups for the digital calibration, as was the case for the paper concurrent calibration.

While this digital calibration model estimated parameters for all digitalPIRLS items, it also re-estimated all paper item parameters. This ensured that the paper and digital item parameters were on the same metric. This also allowed for comparing the digital item parameters with the re-estimated paper item parameters to evaluate differences in item functioning across the paper and digital forms (see Chapter 12).

The digital item parameters were used for estimating ePIRLS items (described in the next section) and also used to estimate student proficiency in all participating countries and benchmarking entities in digitalPIRLS. The estimated digitalPIRLS item parameters are presented in Appendix 11B, along with the re-estimated paperPIRLS item parameters.

ePIRLS Item Calibration

This calibration estimated ePIRLS item parameters on the same metric as the digitalPIRLS items. It included the full digital samples from the 16 digitalPIRLS trend calibration countries. The model estimated ePIRLS item parameters by fixing the digitalPIRLS item parameters estimated from the

digital calibration. Linking ePIRLS to digitalPIRLS also relied on sample equivalence based on the spiraling of booklets within classes.

Exhibit 11.9 shows the number of items included in the ePIRLS calibration by item type and subdomain. There were 91 ePIRLS items and 272 digitalPIRLS regular items.

Exhibit 11.9: **Items for the PIRLS 2021 ePIRLS Calibration**

Item Type	Points	digitalPIRLS Items		ePIRLS Items		Total	
		Items	Points	Items	Points	Items	Points
Multiple Choice	1	123	123	40	40	163	163
Constructed Response	1	91	91	33	33	124	124
	2	50	100	18	36	68	136
	3	8	24	0	0	8	24
Total		**272**	**338**	**91**	**109**	**363**	**447**

Items by Reading Purpose and Comprehension Process

Reading Purpose	digitalPIRLS Items		ePIRLS Items		Total	
	Items	Points	Items	Points	Items	Points
Literary Experience	141	175	0	0	141	175
Acquire and Use Information	131	163	91	109	222	272

Comprehension Process	digitalPIRLS Items		ePIRLS Items		Total	
	Items	Points	Items	Points	Items	Points
Retrieving and Straightforward Inferencing	165	187	51	58	216	245
Interpreting, Integrating, and Evaluating	107	151	40	51	147	202
Total	**272**	**338**	**91**	**109**	**363**	**447**

Exhibit 11.10 presents the sample sizes for the ePIRLS calibration. From the 16 digital calibration countries, this included the same 61,256 students administered digitalPIRLS booklets from the digital calibration, as well as the 20,440 students and 10,147 students administered ePIRLS and hybrid booklets, respectively.

Exhibit 11.10: Sample Sizes for the PIRLS 2021 ePIRLS Calibration

Country	Item Calibration			Proficiency Estimation		
	Digital	ePIRLS	Hybrid	Digital	ePIRLS	Hybrid
Belgium (Flemish)	3,408	1,145	561	3,408	1,145	561
Chinese Taipei	3,696	1,242	617	3,696	1,242	617
Croatia	—	—	—	2,623	868	446
Czech Republic	4,388	1,488	745	4,388	1,488	745
Denmark	3,220	1,076	525	3,220	1,076	525
Finland	4,676	1,561	781	4,676	1,561	781
Germany	3,068	1,030	513	3,068	1,030	513
Hungary	—	—	—	3,545	1,175	592
Israel	—	—	—	3,267	1,083	540
Italy	3,628	1,211	601	3,628	1,211	601
Kazakhstan	—	—	—	4,668	1,575	780
Lithuania	—	—	—	3,097	1,021	505
Malta	—	—	—	2,050	648	332
New Zealand	3,707	1,233	617	3,707	1,233	617
Norway (5)	3,594	1,192	596	3,594	1,192	596
Portugal	4,117	1,341	653	4,117	1,341	653
Qatar	—	—	—	3,497	1,180	581
Russian Federation	3,488	1,146	583	3,488	1,146	583
Saudi Arabia	—	—	—	3,186	1,054	538
Singapore	4,482	1,490	747	4,482	1,490	747
Slovak Republic	3,220	1,088	533	3,220	1,088	533
Slovenia	3,416	1,139	555	3,416	1,139	555
Spain	5,691	1,910	950	5,691	1,910	950
Sweden	3,457	1,148	570	3,457	1,148	570
United Arab Emirates	—	—	—	18,322	6,085	3,041
Benchmarking Participants						
Alberta, Canada	—	—	—	2,006	682	332
British Columbia, Canada	—	—	—	3,111	1,052	512
Newfoundland & Labrador, Can.	—	—	—	1,623	542	280
Quebec, Canada	—	—	—	2,496	820	423
Moscow City, Russian Fed.	—	—	—	3,837	1,270	638
Abu Dhabi, UAE	—	—	—	6,941	2,302	1,138
Dubai, UAE	—	—	—	5,137	1,705	869
TOTAL	**61,256**	**20,440**	**10,147**	**130,662**	**43,502**	**21,694**

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

The ePIRLS item parameters estimated from this calibration are presented in Appendix 11B.

Phase 2: Principal Component Analysis

The second phase of the PIRLS 2021 achievement scaling involved creating principal components for use in conditioning. Conditioning refers to applying a latent regression model that includes all available contextual information to improve the statistical properties of the estimated student proficiency plausible values. Ideally, all student-level contextual data would be included in the conditioning model, but because PIRLS has so many 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 component analysis (PCA) 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. Unique dummy coded variables were created for "not administered," "not applicable," and "omitted" responses, respectively.

- Context variables with numerous response options (such as year of birth) were recoded using criterion scaling (Beaton, 1969). 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 student-level average of the reading EAP scores produced by the item calibrations.

- Separately for each country, all the dummy-coded and criterion-scaled variables were included in a principal component analysis. Those principal components accounting for 90 percent of the variance of all context variables were retained for use as conditioning variables. Because the principal component analysis was performed separately for each country and benchmarking entity, different numbers of principal components were required to account for 90 percent of the common variance in each country's context variables. In addition, the number of principal components retained was limited to no more than 5 percent of a country's student sample size to avoid over-specification of the conditioning model. This constraint played a major role with the digitalPIRLS bridge samples due to their smaller sample size. As a result, principal components were selected that either retained 90 percent of common variance or 5 percent of the sample size, whichever led to fewer principal components.

In addition to the principal components, students' 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 latent regression conditioning model.

Exhibits 11.11 and 11.12 provide details on the conditioning models used for proficiency estimation of the paperPIRLS data and bridge data, respectively. In these exhibits, conditioning model details also are reported for the PIRLS 2016 data. Exhibit 11.13 provides details on the conditioning models used for proficiency estimation of the digitalPIRLS (including ePIRLS) data.

Exhibit 11.11: **Conditioning Model Dimensionality for the paperPIRLS 2021 Data**

Country	PIRLS 2016				PIRLS 2021			
	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
Albania	—	—	—	—	2	577	210	76
Australia	2	543	277	90	2	287	163	90
Austria	2	548	218	79	2	582	240	81
Azerbaijan	3	537	299	89	3	575	260	82
Bahrain	3	549	274	85	3	580	260	82
Belgium (French)	2	530	231	82	2	565	213	78
Brazil	—	—	—	—	2	571	247	82
Bulgaria	2	533	214	81	2	565	202	78
Cyprus	—	—	—	—	3	583	229	78
Egypt	—	—	—	—	2	583	325	90
England	2	248	146	90	2	279	162	90
France	2	547	238	82	2	581	266	84
Georgia	3	549	287	88	2	581	262	83
Hong Kong SAR	2	545	167	71	2	583	191	74
Iran, Islamic Rep. of	2	549	219	80	2	583	298	88
Ireland	2	549	230	82	2	582	233	81
Jordan	—	—	—	—	2	583	307	89
Kosovo	—	—	—	—	2	577	227	78
Latvia	3	549	207	77	3	583	218	78
Macao SAR	4	549	202	75	4	583	254	82

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

Country	PIRLS 2016				PIRLS 2021			
	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
Montenegro	—	—	—	—	2	579	224	77
Morocco	2	549	274	86	2	583	324	90
Netherlands	2	543	210	81	2	562	215	80
North Macedonia	—	—	—	—	3	582	146	66
Northern Ireland	2	511	184	79	2	539	202	80
Oman	3	549	319	90	3	583	266	82
Poland	2	536	220	81	2	574	208	77
Serbia	—	—	—	—	2	575	201	76
South Africa	—	—	—	—	12	565	335	90
Turkiye	—	—	—	—	2	551	301	90
Uzbekistan	—	—	—	—	4	576	292	86
Benchmarking Participant								
South Africa (6)	—	—	—	—	3	565	328	90

Exhibit 11.12: Conditioning Model Dimensionality for the PIRLS 2021 Bridge Data

Country	PIRLS 2016				PIRLS 2021			
	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
Belgium (Flemish)	2	545	259	86	2	576	81	52
Chinese Taipei	2	546	216	79	2	572	83	54
Croatia	—	—	—	—	2	578	61	49
Czech Republic	2	540	276	88	2	569	95	58
Denmark	2	549	175	72	2	583	70	48
Finland	3	548	244	84	3	582	103	59
Germany	2	545	197	77	2	577	67	54
Hungary	2	525	231	82	2	558	84	56
Israel	3	511	202	78	3	535	89	57
Italy	2	543	197	74	2	571	98	53
Kazakhstan	3	531	246	82	3	583	160	68
Lithuania	4	528	215	80	4	564	75	56
Malta	—	—	—	—	3	558	41	41
New Zealand	7	519	277	90	8	578	111	64
Norway (5)	3	530	211	78	3	560	83	53
Portugal	2	548	232	81	2	583	104	56
Qatar	3	546	307	90	3	581	67	51
Russian Federation	2	525	228	81	2	559	109	59
Saudi Arabia	3	549	237	79	3	583	93	56
Singapore	2	549	300	90	2	563	99	59
Slovak Republic	3	549	272	87	3	583	82	55
Slovenia	2	544	224	81	2	578	70	50
Spain	8	542	302	90	6	572	78	52
Sweden	2	525	226	82	2	551	93	59
United Arab Emirates	6	549	315	90	5	583	99	59
United States	9	250	147	90	8	265	82	67
Benchmarking Participant								
Moscow City, Russian Fed.	2	525	214	78	2	551	84	51

Exhibit 11.13: Conditioning Model Dimensionality for the digitalPIRLS 2021 Data

Country	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Belgium (Flemish)	2	576	255	84
Chinese Taipei	2	575	277	87
Croatia	2	578	196	76
Czech Republic	2	573	303	90
Denmark	2	583	241	81
Finland	3	582	305	90
Germany	2	577	230	82
Hungary	2	558	265	86
Israel	3	537	244	84
Italy	2	571	272	84
Kazakhstan	3	583	323	90
Lithuania	4	566	231	83
Malta	3	561	151	67
New Zealand	8	578	277	88
Norway (5)	3	561	269	86
Portugal	2	583	305	88
Qatar	3	583	262	84
Russian Federation	2	559	260	84
Saudi Arabia	3	583	238	80
Singapore	2	564	299	90
Slovak Republic	3	583	242	83
Slovenia	2	579	255	84
Spain	6	572	310	90
Sweden	2	551	258	86
United Arab Emirates	5	583	321	90
Benchmarking Participants				
Alberta, Canada	3	564	151	70
British Columbia, Canada	3	565	233	83
Newfoundland & Labrador, Can.	2	562	122	64
Quebec, Canada	3	565	186	74
Moscow City, Russian Fed.	2	551	287	88
Abu Dhabi, UAE	3	583	316	90
Dubai, UAE	3	583	303	90

Phase 3: Latent Regression Population Modeling

Educational Testing Service's MGROUP programs (Rogers et al., 2006; Sheehan, 1985) were used to estimate the latent regression model and to generate plausible values for the PIRLS 2021 paper and digital data. These programs take as input the students' responses to the items they were given, the item parameters estimated at the calibration stage, and the conditioning variables, and generate as output the estimated regression effects and the residual variance-covariance matrix, as well as plausible values that represent the posterior distribution of student proficiency given their achievement and contextual data (e.g., Mislevy, 1991; Thomas, 1993; von Davier et al., 2006; von Davier & Sinharay, 2013). More details on the latent regression model are available in Chapter 10.

Certain versions of the MGROUP set of programs allow multi-dimensional latent regression models 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 modeling feature implemented in MGROUP was used to generate plausible values for the PIRLS 2021 reading purpose and comprehension process domains. The estimation of these plausible values for the reading subdomains relied on multi-dimensional IRT models using the item parameters estimated for the overall reading scale, as well as the same set of conditioning variables. PIRLS 2021 used two two-dimensional scaling models, one to estimate proficiency scores for the two reading purposes and a second for the two comprehension processes.

The latent regression models for overall reading and for the subdomains were estimated for the data from all PIRLS 2021 countries and benchmarking participants. First, plausible values were generated for paperPIRLS and bridge data, using the item parameters from the paper concurrent calibration and the conditioning variables. Plausible values were generated for the digitalPIRLS data using the item parameters from the digital and ePIRLS calibrations.

The paper item parameters also were used to generate plausible values for the PIRLS 2016 assessment data for all trend countries. These additional plausible values were used to establish the linear transformation necessary to place the paperPIRLS and bridge 2021 data on the PIRLS trend scale.

Phase 4: Scale Transformation

To provide results for the PIRLS 2021 assessments on the existing PIRLS achievement scale, the 2021 plausible values had to be transformed to the PIRLS reporting metric. This process involved performing a series of linear transformations on the plausible values derived from the paperPIRLS and digitalPIRLS data. PIRLS 2021 yielded two sets of transformation constants. The first set put the paper-based 2021 results on the PIRLS trend scale. It was derived from the published and re-estimated PIRLS 2016 ability distributions. The second set of scale transformation constants put the digital results on the PIRLS trend scale, and was obtained from the transformed

bridge distribution and the untransformed digital ability distributions. The equations used to derive transformation constants and to transform plausible values are provided in Chapter 10.

paperPIRLS Scale Transformation

The first paperPIRLS scale transformation put all paperPIRLS 2021 results, including bridge data, on the PIRLS trend scale by aligning the re-estimated PIRLS 2016 ability distribution with the published PIRLS 2016 ability distribution. Five pairs of transformation constants were obtained— one for each plausible value of overall reading proficiency. First, international means and standard deviations were computed for the trend calibration countries using the published 2016 plausible values based on the 2016 item calibration. Next, means and standard deviations were calculated using the re-scaled 2016 plausible values based on the 2021 paper concurrent calibration model. The 27 trend calibration countries included in the paper concurrent calibration contributed equally to the calculation of these transformation constants.

Exhibit 11.14 shows the PIRLS 2021 transformation constants for the paperPIRLS 2021 data, including paperPIRLS countries' data and the bridge data collected in digitalPIRLS countries. These transformation constants were applied to overall reading, the reading purpose subdomains, and the comprehension process subdomains.

Exhibit 11.14: Scale Transformation Constants for the PIRLS 2021 Paper Data

	PIRLS 2016 Published Scores		PIRLS 2016 Re-Scaled Scores		A_{ik}	B_{ik}
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	538.13388	79.45850	0.27954	0.78075	509.68449	101.77257
PV2	537.64200	80.04106	0.28208	0.77943	508.67491	102.69179
PV3	536.99064	80.27275	0.28314	0.78015	507.85754	102.89396
PV4	537.35784	80.23786	0.28233	0.78174	508.37981	102.63957
PV5	537.35481	79.92416	0.28250	0.77923	508.37960	102.56856

Under this approach to measuring trends, achievement changes over time in the reading purpose and comprehension process subdomains are within the context of achievement in overall reading. Trends are not established separately for each purpose or process. The subscales are based on only half the assessment items, making them somewhat less robust than the overall reading achievement results based on the entire scale. The subscale trend results are not included in the *PIRLS 2021 International Results in Reading* report because they are less stable than the overall trend results and could be even further influenced by COVID-19 to an unknown degree.

The paperPIRLS transformation constants also were applied to the re-scaled 2016 plausible values in order to evaluate the variation of the linking adjustment between 2021 and 2016 scaling models (see results in the later "Validating the PIRLS 2021 Achievement Results" section).

digitalPIRLS Scale Transformation

The digitalPIRLS linear scale transformation aligned the pooled digital ability distribution from the digital item calibration with the transformed bridge ability distribution resulting from the paper concurrent calibration. The joint scaling of bridge and digital accomplished the linking and put the data on the same scale. This additional transformation eliminates any remaining sample differences between bridge and digital sample distributions. The 16 digital countries included in the paper and digital item calibrations, with their randomly equivalent bridge and digital samples, were the basis for calculating the transformation constants.

Transformation constants were obtained by first computing the international means and standard deviations of the bridge plausible values for overall reading, which already had been placed on the PIRLS trend metric in the previous transformation step. Next, the means and standard deviations of the digital plausible values were calculated based on the digital item calibration. These results were treated as interim since they were only based on students administered one of the 18 digitalPIRLS booklets. The linear transformation constants were determined by aligning the means and standard deviations of the untransformed digital plausible values with the means and standard deviations of the transformed bridge plausible values.

Exhibit 11.15 shows the five pairs of transformation constants for the digitalPIRLS overall reading scale, one for each plausible value. These transformation constants were applied to all PIRLS 2021 digital data for overall reading as well as for the reading purposes and comprehension processes. The transformed plausible values of the full digital samples, including the plausible values from hybrid and ePIRLS booklets, were used for reporting achievement for digitalPIRLS countries. The transformed interim results from the digital calibration, based on the 18 digitalPIRLS booklets, were used for validation purposes only (see next section).

	Transformed Bridge Scores		Untransformed digitalPIRLS Scores		A_{ik}	B_{ik}
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	537.91609	76.65646	-0.00161	0.76091	538.07815	100.74315
PV2	537.01898	77.26746	-0.00116	0.76181	537.13672	101.42598
PV3	536.13574	77.07134	-0.00172	0.76399	536.30970	100.88021
PV4	536.51431	77.32874	-0.00055	0.76019	536.57072	101.72353
PV5	536.81941	77.26436	-0.00113	0.76305	536.93362	101.25662

Validating the PIRLS 2021 Achievement Results

Scaling the PIRLS 2021 achievement data included extensive analyses throughout the process to ensure the quality of the results. In this section, three major aspects of the scaling are addressed:

- Evaluating item fit to the PIRLS 2021 assessment data
- Examining the variation in the trend linking error across countries
- Evaluating the integration of ePIRLS with digitalPIRLS

Another important aspect of the achievement scaling involved validating the randomly equivalent samples linking approach by comparing the bridge and digital achievement results. These analyses are described in Chapter 12.

Evaluating Item Fit to the PIRLS 2021 Assessment Data

To ensure reliable and accurate item parameter estimates of the PIRLS 2021 achievement data, a series of IRT-based checks were performed during the item calibration phase. Several methods were used to verify the fit of the IRT model item parameters to the data. These included examining graphical displays of item characteristic curves (ICCs) to check the empirical and fitted item response functions and to compare the empirical curves for trend items between 2021 and 2016 cycles. In addition, quantitative inspections were conducted with the root mean square difference (RMSD) statistic. Furthermore, comparisons of item parameters were conducted between PIRLS 2016 and 2021, as well as between the two sets of paperPIRLS 2021 item parameters estimated from the paper concurrent calibration and digital calibration.

Item Characteristic Curves

Item fit was assessed by visually comparing the item response function curves generated using the item parameters estimated from the data with the empirical item response function curves calculated from the latent abilities estimated for each student that responded to an item. The empirical functions are themselves based on an estimated latent ability distribution that uses the IRT model and therefore also are referred to as item functions based on *pseudo counts*. When the empirical results for an item fall near the fitted curves, the IRT model for that item fits the data well and provides an accurate and reliable measurement of the underlying proficiency scale.

Plots of these response function curves are called item characteristic curves (ICC). ICC plots were examined for all paperPIRLS items, digitalPIRLS items, and ePIRLS items. The plots in Exhibits 11.16 and 11.17 show examples of the empirical and fitted item response functions for dichotomously scored (correct/incorrect) multiple-choice and constructed-response items, respectively. In each plot, the horizontal axis represents the proficiency scale on the logit metric, 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, with the item slope parameter represented by the slope of the curve between the two inflexion points, the difficulty or location parameter represented by the point on the horizontal axis where the probability of a correct response is 50 percent, and, for multiple-choice items, a lower asymptote corresponding to the guessing parameter.

Empirical results based on *pseudo counts* are represented by circles. The empirical results are obtained by first dividing the logit proficiency scale into intervals of equal size and then counting the number of students responding to the item whose *estimated* latent abilities (estimated EAP scores) 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 pseudo count (the estimated number of students) contributing to the empirical proportion correct in its corresponding interval.

Exhibit 11.16: Example Item Response Function for a Dichotomous Multiple-Choice Item from PIRLS 2021

Probability of a Correct Response for Estimated Proficiency (With Information Function)
PIRLS 2021 – CLB
Item = RE31M07 a = 1.286 b = −0.642 c = 0.17
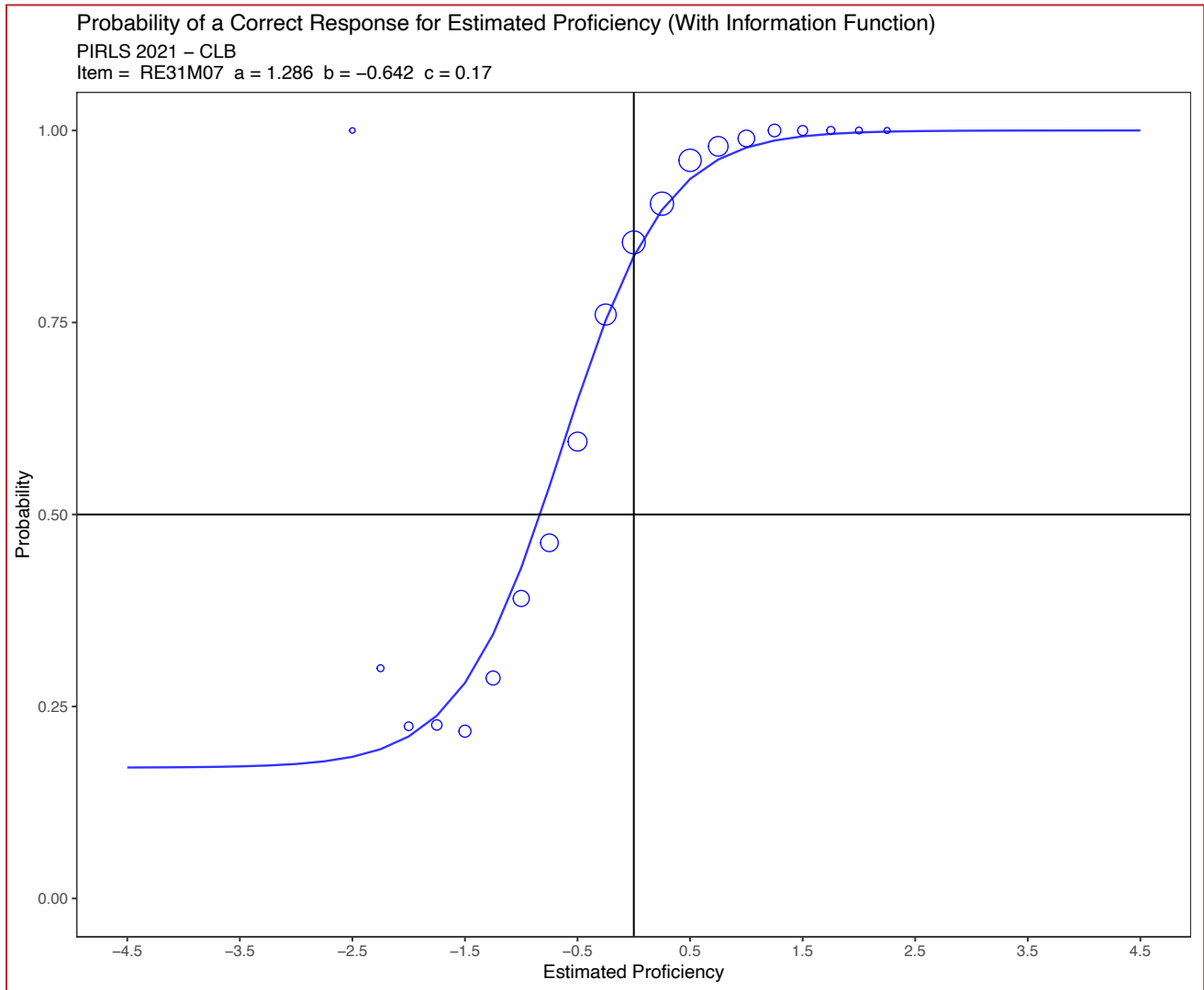
Although a single set of item parameters was estimated for any given item in the concurrent calibration for 2021, trend items have two empirical curves, one for the data from each assessment cycle. Plotting both empirical curves from 2021 and 2016 allowed for a visual inspection of the invariance of the item parameters between cycles, a key aspect of the link to the trend scale. Exhibit 11.17 shows the ICC for a paperPIRLS 2021 trend item, with its single fitted curve and two empirical curves: the red bubbles represent the empirical curve based on the PIRLS 2016 response data, the blue curve the empirical curve based on the PIRLS 2021 response data.

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

CHAPTER 11: IMPLEMENTING THE SCALING METHODOLOGY
METHODS AND PROCEDURES: PIRLS 2021 TECHNICAL REPORT   11.32

Probability of a Correct Response for Estimated Proficiency (With Information Function)
PIRLS 2021 – CLB
Item = RP31W11 a = 1.367 b = 0.452

The ICC plot in Exhibit 11.18 shows the empirical and fitted item response functions for a polytomous item (scored 0, 1, or 2). As for the dichotomous item plots above, the horizontal axis represents the proficiency scale in logits, but in this example the vertical axis represents the probability of a response in a certain category. The fitted curves based on the estimated item parameters are shown as solid lines and the empirical results are represented by circles. The interpretation of the circles is the same as in Exhibit 11.16. The curve starting at the top left of the chart plots the probability of a score of zero on the item. This probability should always decrease as proficiency increases. The bell-shaped curve shows the probability of a score of one point—partial credit, which should start low approaching zero for low-ability students, reaching a maximum for

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

CHAPTER 11: IMPLEMENTING THE SCALING METHODOLOGY
METHODS AND PROCEDURES: PIRLS 2021 TECHNICAL REPORT   11.33

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 11.18: **Example Item Response Function for a Polytomous Constructed-Response Item from PIRLS 2021**

Probability of a Correct Response for Estimated Proficiency (With Information Function)
PIRLS 2021 – CLB
Item = RE41O10 a = 0.926 b = 0.098 step1 = 0.194 step2 = −0.194
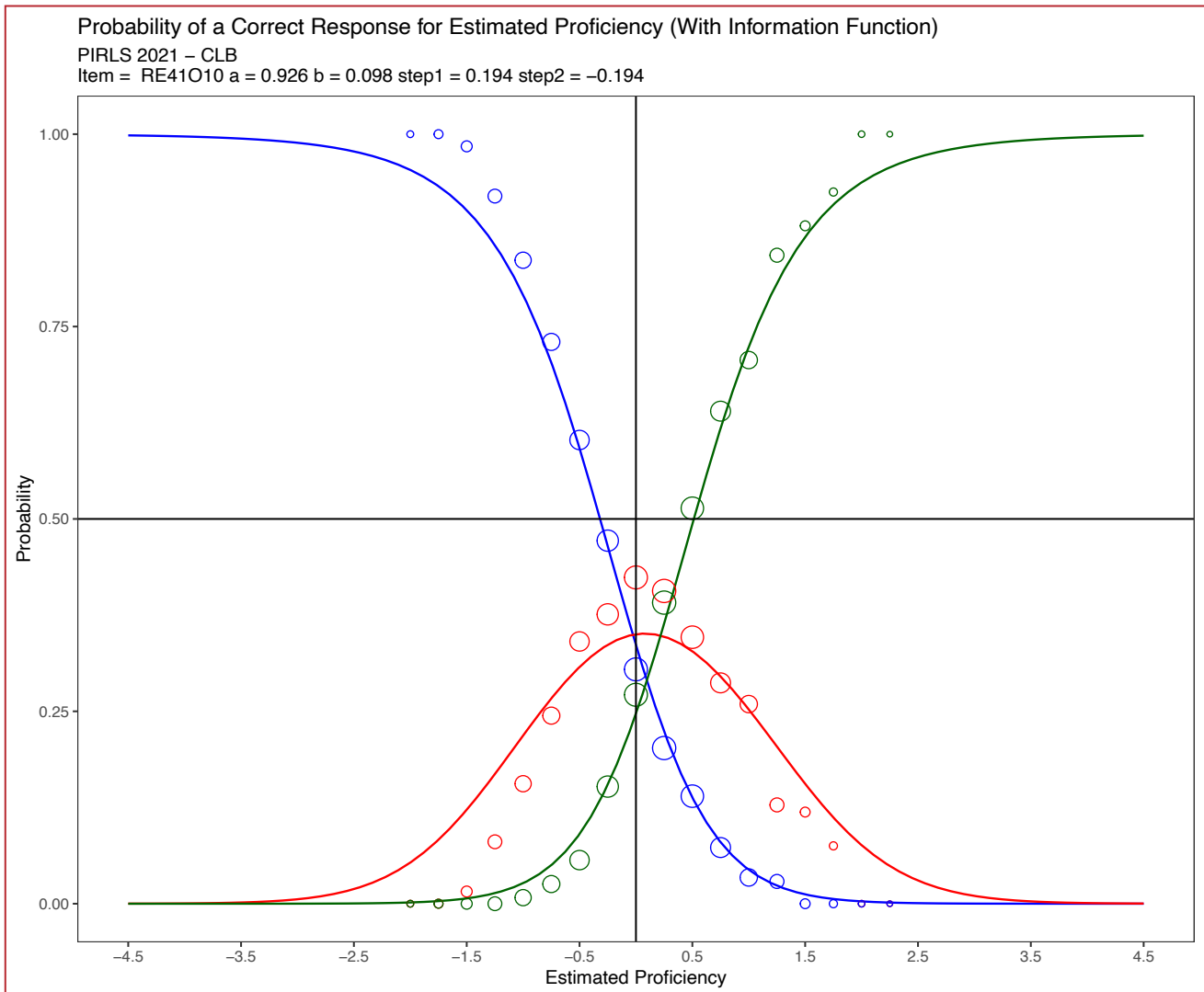
Root Mean Square Difference (RMSD)

In addition to graphical model fit assessment, item fit also was checked using the root mean square difference (RMSD) statistic. The RMSD is the square root of the average of squared differences (i.e., the area) between the empirical curve, shown as bubbles in the ICCs above, and the fitted curve, weighted by the number of students at each ability interval.

RMSD values were computed for all PIRLS 2021 items, and are reported in the item parameter tables in Appendices 11A and 11B. They are also presented graphically in Exhibits 11.19 and 11.20. In each exhibit, the items are sorted from smallest to largest RMSD values. For trend items with two RMSD values, the largest of the two determined the order. All items in the PIRLS 2021 achievement scaling had good RMSD fit statistics.

Misfitting items were identified by using the median absolute deviation (MAD) outlier detection method on the RMSD statistic. MAD is a robust measure of dispersion which was used as a flagging rule instead of an arbitrary cut-off value (von Davier & Bezirhan, 2022). This method flags an item as a possible misfit if its distance from the median of the absolute distances of all other observations exceeds a predetermined threshold. For the PIRLS 2021 achievement scaling, a threshold of 3 was used to identify items that needed further evaluation or possible deletion.

Exhibit 11.19: **RMSD Statistics for Items in the PIRLS 2021 Paper Concurrent Calibration**
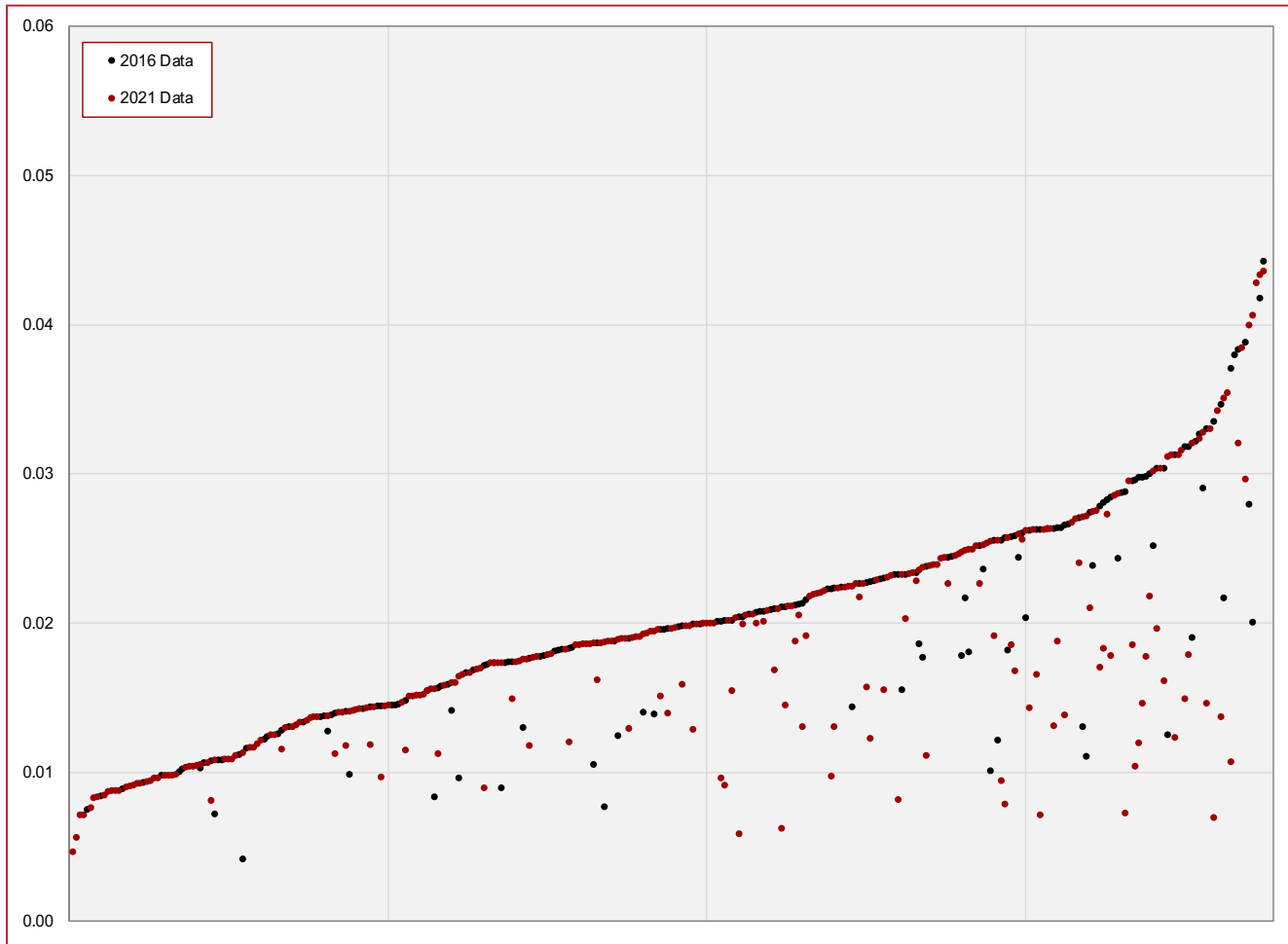
Exhibit 11.20: RMSD Statistics for Items in the PIRLS 2021 Digital and ePIRLS Calibrations
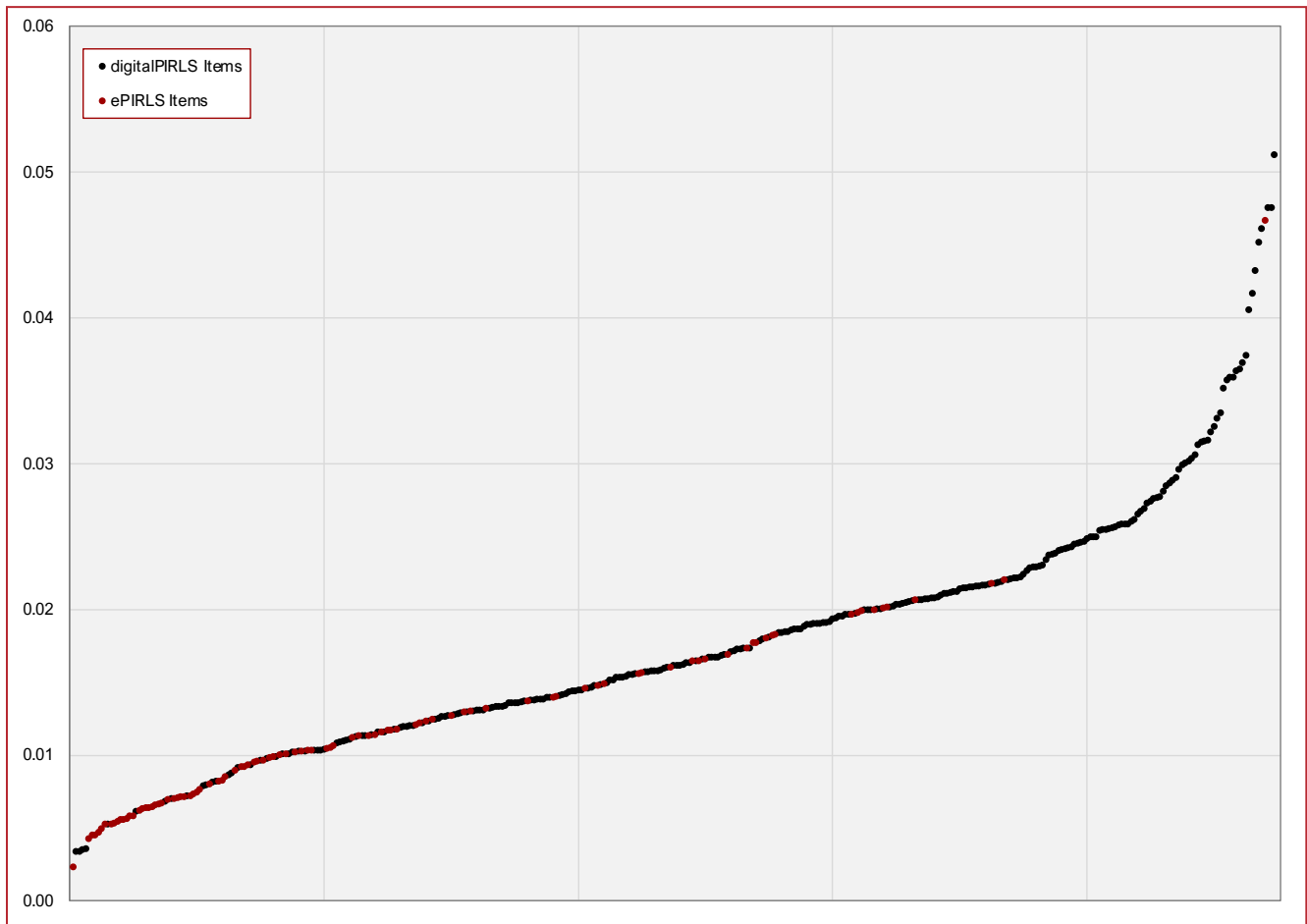
Item Parameter Comparisons

Another step in evaluating the paper concurrent calibration consisted of a comparison of 2016 item parameters re-estimated from the 2021 paper concurrent calibration against their parameter estimates from the 2016 concurrent calibration. Exhibits 11.21 through 11.23 show graphically the comparisons of item parameter estimates for difficulty, discrimination, and guessing, respectively. A high degree of agreement was observed, with correlation coefficients ranging between 0.89 and 0.99.

Exhibit 11.21: Plot Comparing Item Difficulty Parameters between PIRLS 2016 and PIRLS 2021 Calibration Models
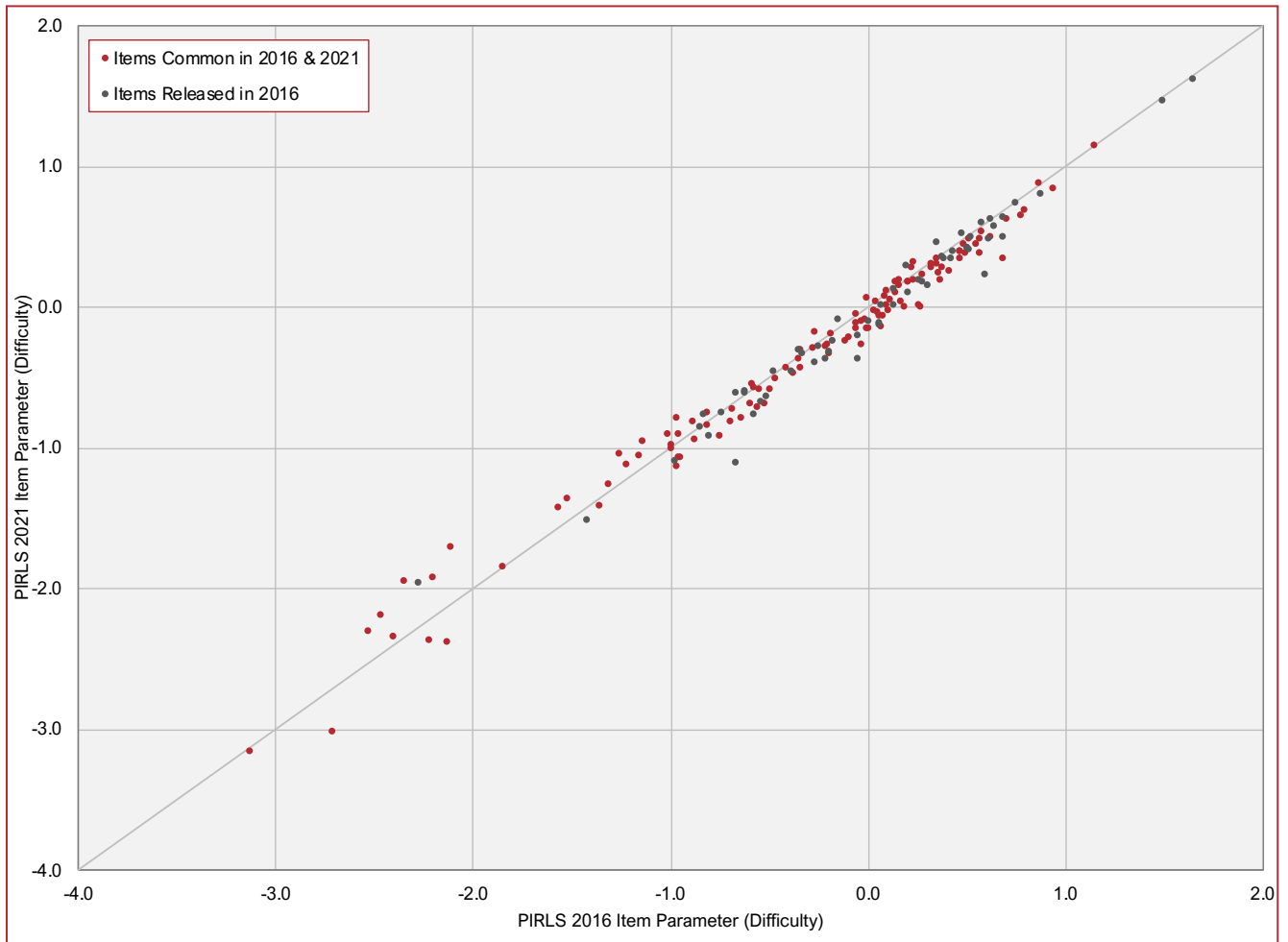
Exhibit 11.22: Plot Comparing Item Discrimination Parameters between PIRLS 2016 and PIRLS 2021 Calibration Models

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

CHAPTER 11: IMPLEMENTING THE SCALING METHODOLOGY
METHODS AND PROCEDURES: PIRLS 2021 TECHNICAL REPORT 11.38
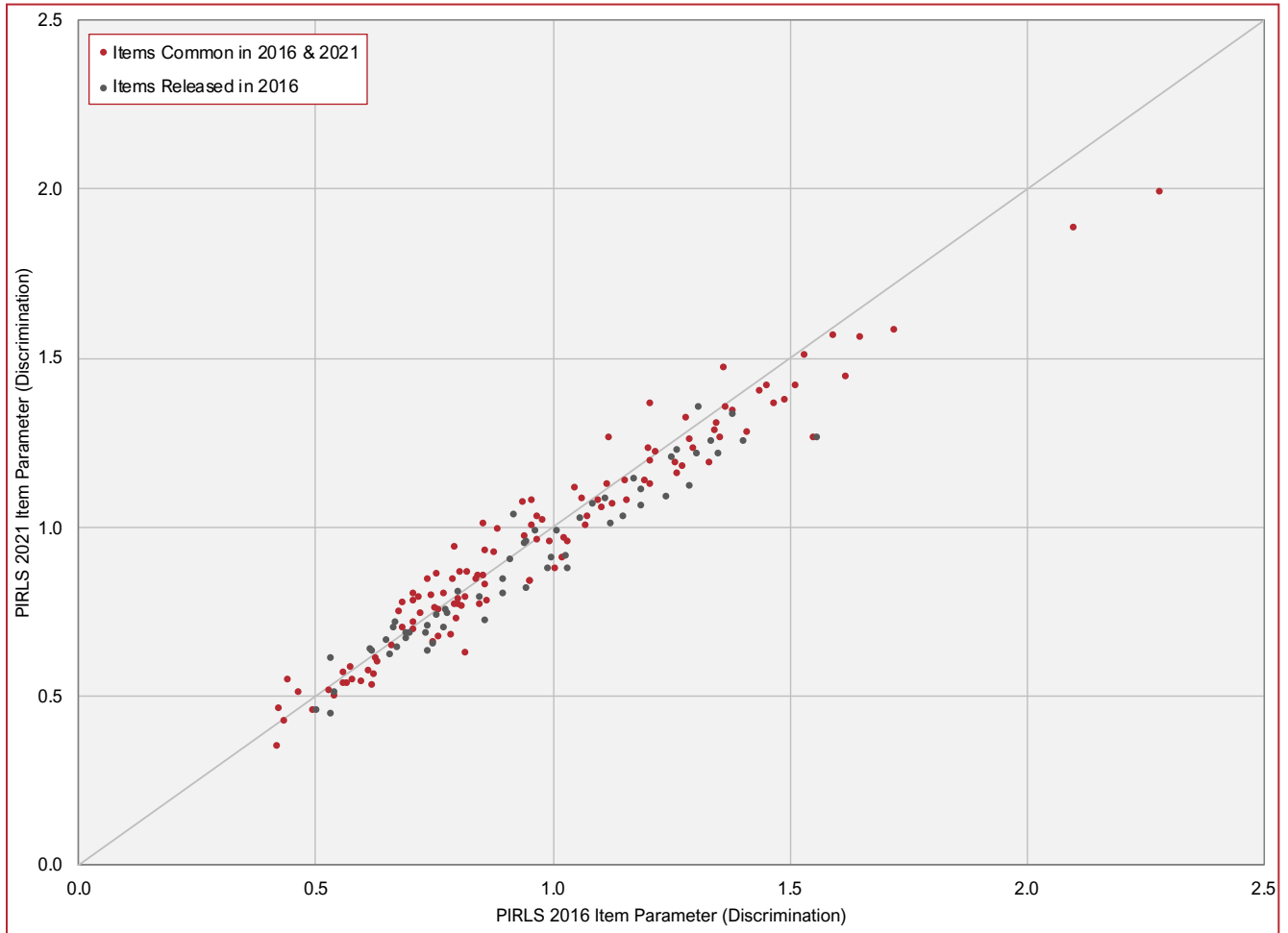
Exhibit 11.23: **Plot Comparing Item Guessing Parameters between PIRLS 2016 and PIRLS 2021 Calibration Models**
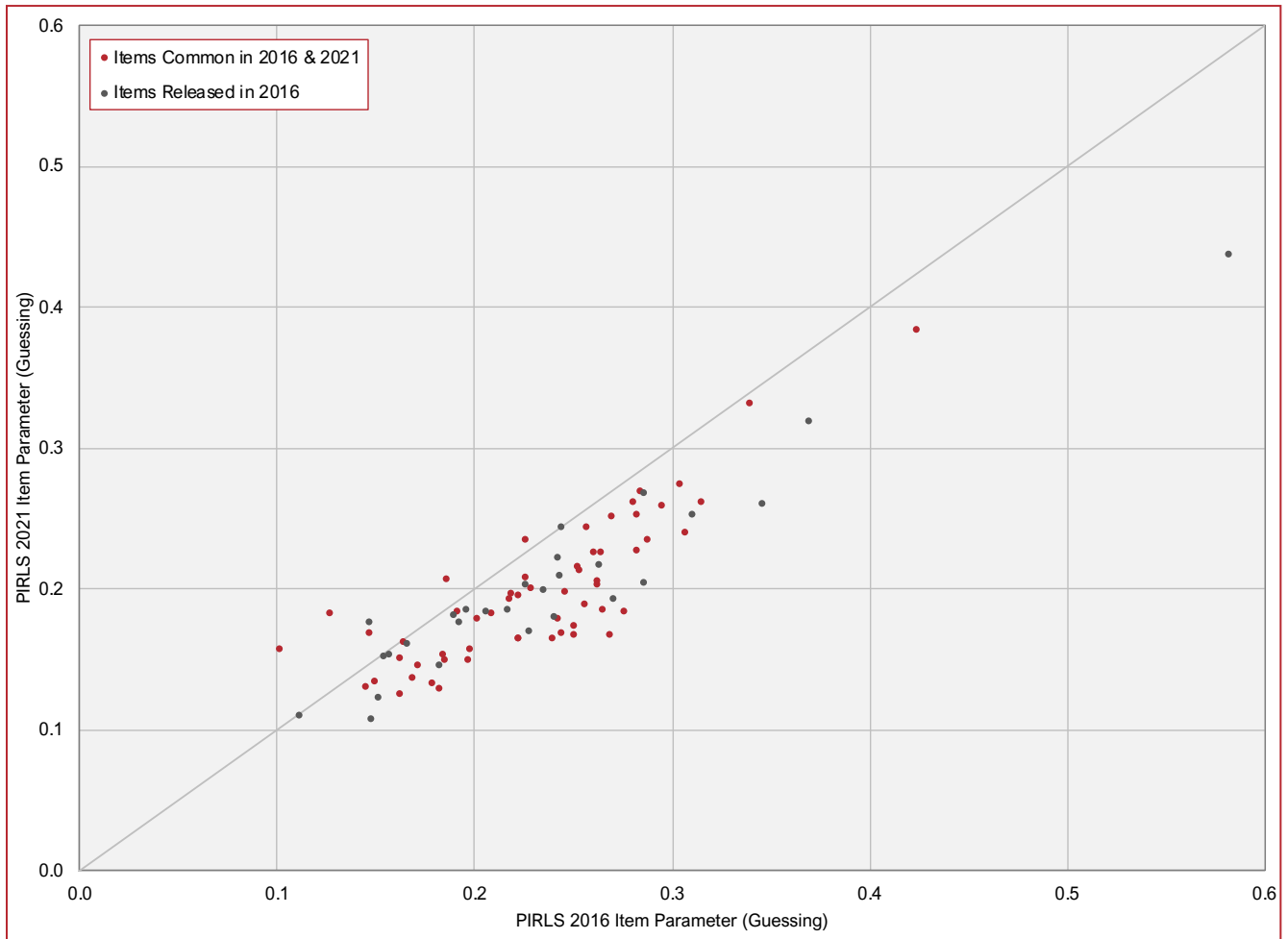
Comparisons were conducted also between the re-estimated paper item parameters from the digital calibration and the paper item parameter estimates from the paper concurrent calibration. Exhibits 11.24 through 11.26 display the comparison of the two sets of paper item parameter estimates for difficulty, discrimination, and guessing, respectively. The plots show a high level of agreement with almost perfect correlations between the two sets of item parameters.

The small scale shift between the difficulty parameters (Exhibit 11.24) is due to the fact that the digital data were added to the paper data in the digital calibration. With both paper and digital item parameters freely estimated in the digital calibration, different identification constraints resulting from adding the digital data led to an offset. This offset is accounted for in the linking.

Exhibit 11.24: Plot Comparing Paper Item Difficulty Parameters between Paper and Digital Calibration Models
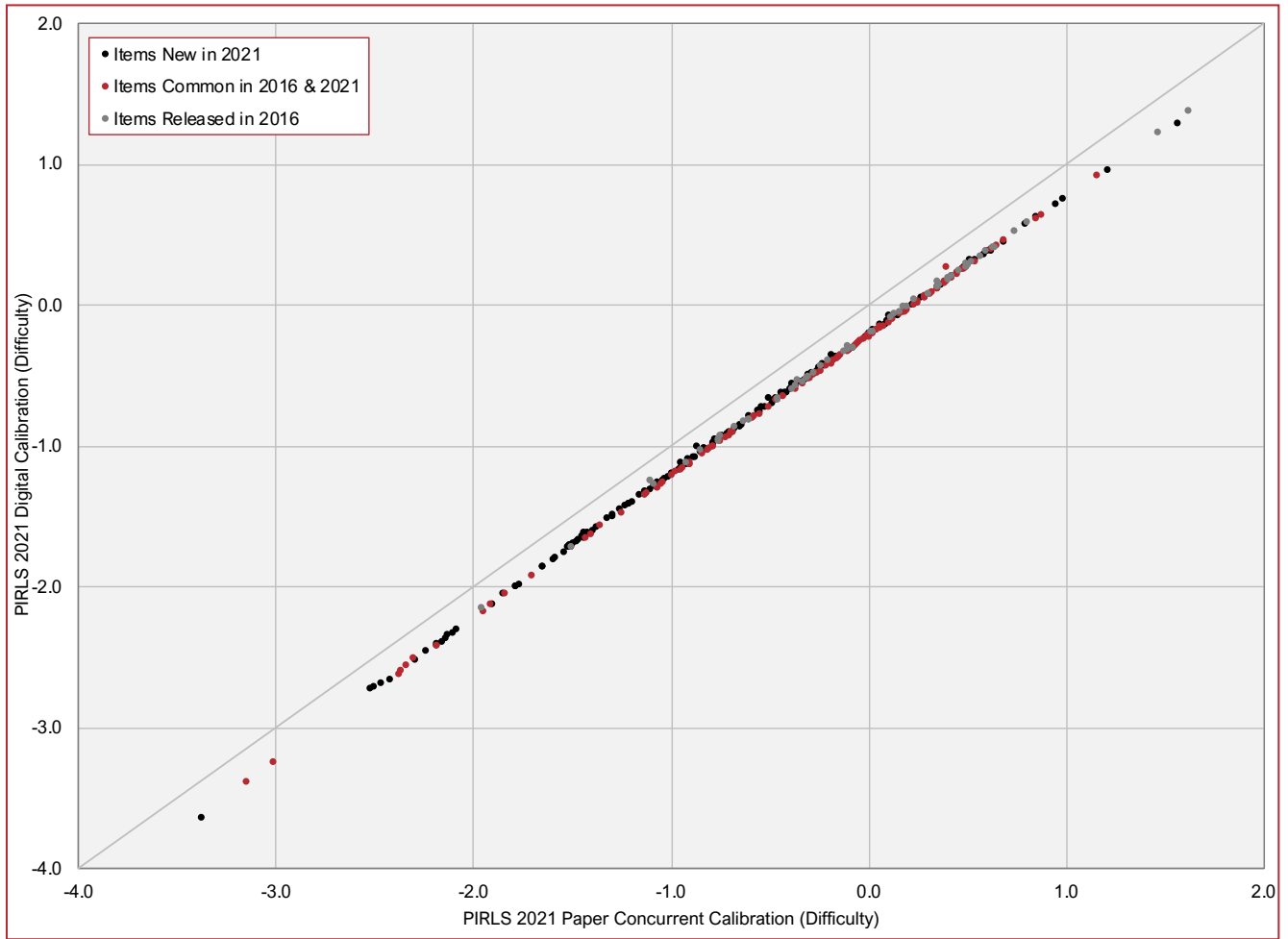
Exhibit 11.25: Plot Comparing Paper Item Discrimination Parameters between Paper and Digital Calibration Models

Examining the Variation in Trend Linking Error

A key aspect of reporting the PIRLS 2021 results on the PIRLS trend scale is the ability to accurately re-estimate the PIRLS 2016 achievement results based on a concurrent calibration of the 2016 and 2021 data across a common set of trend countries. As described earlier, this re-estimation serves to establish the linear transformation that places the PIRLS 2021 results on the PIRLS trend scale. Although this transformation is set globally to match the overall mean and standard deviation across the trend countries, it should also achieve an adequate alignment of the 2016 results between calibrations for each individual trend country. The difference between a trend country's PIRLS 2016 achievement mean published back in 2016 and re-estimated in 2021 provides a measure of the quality of the link between the two assessments.

Exhibit 11.27 provides results on the trend linking error associated with the paper-based PIRLS 2021 results. Results are included for trend calibration countries, as well as trend countries that did not contribute to setting the scale. There was a good agreement between the countries' published and re-estimated 2016 results. Aggregated across all countries, there is no difference at the international level. Although there are small differences at the country level, the differences are within two points and the standard errors rarely exceed 2 points—both for calibration countries and those that did not contribute. These small difference are expected, given that the PIRLS 2016 data were calibrated under two different calibration models with different sets of countries and items.

It is noteworthy that the relatively larger linking differences such as 4 points or more mainly occurred in the relatively lower performing countries (e.g., mean score below 500) as shown in the exhibit. This is mainly due to a relatively larger percent of omitted and not-reached responses in these countries that were treated with the strength-of-evidence approach for the 2021-2016 concurrent calibration, but were previously treated with the deterministic approach to treating item non-response in the 2016-2011 concurrent calibration.

Exhibit 11.27: **Trend Linking Error Variance between PIRLS 2016 and PIRLS 2021 Calibrations**

Country		PIRLS 2016 Published Results		PIRLS 2016 Re-Scaled Results		Difference
		Mean	Standard Deviation	Mean	Standard Deviation	
Australia	*	544 (2.5)	84 (1.6)	544 (2.6)	86 (1.9)	0 (1.2)
Austria	*	541 (2.4)	65 (1.4)	541 (2.6)	64 (1.4)	0 (1.2)
Azerbaijan	*	472 (4.2)	86 (2.8)	477 (4.1)	79 (2.5)	4 (1.1)
Bahrain		446 (2.3)	98 (1.5)	443 (2.4)	97 (1.7)	-3 (0.9)
Belgium (Flemish)	*	525 (1.9)	61 (0.9)	522 (2.0)	62 (1.2)	-3 (0.6)
Belgium (French)	*	497 (2.6)	69 (1.3)	498 (2.3)	66 (1.3)	1 (0.8)
Bulgaria	*	552 (4.2)	85 (2.7)	554 (4.0)	84 (1.7)	2 (1.7)
Chinese Taipei	*	559 (2.0)	64 (1.0)	560 (2.1)	66 (1.0)	1 (1.1)
Czech Republic	*	543 (2.1)	68 (2.0)	545 (2.0)	65 (1.6)	1 (1.0)
Denmark	*	547 (2.1)	68 (1.5)	547 (2.2)	66 (1.3)	-1 (1.4)
England		559 (1.9)	79 (1.2)	558 (2.0)	81 (1.2)	-1 (0.9)
Finland	*	566 (1.8)	67 (1.6)	566 (1.9)	66 (1.3)	0 (0.9)
France	*	511 (2.2)	69 (1.4)	513 (2.2)	66 (1.1)	2 (0.8)
Georgia		488 (2.8)	79 (1.6)	492 (2.8)	72 (1.4)	4 (1.3)
Germany	*	537 (3.2)	78 (3.2)	539 (2.9)	74 (2.3)	2 (0.9)
Hong Kong SAR	*	569 (2.7)	64 (1.5)	569 (2.6)	62 (1.3)	1 (0.9)

Country		PIRLS 2016 Published Results		PIRLS 2016 Re-Scaled Results		Difference
		Mean	Standard Deviation	Mean	Standard Deviation	
Hungary		554 (2.9)	75 (1.6)	554 (3.0)	75 (1.5)	0 (1.0)
Iran, Islamic Rep. of		431 (3.9)	103 (3.2)	437 (3.3)	93 (2.4)	6 (1.8)
Ireland		567 (2.5)	74 (1.5)	566 (2.7)	76 (1.4)	0 (0.9)
Israel		530 (2.5)	90 (1.8)	532 (2.5)	87 (1.9)	2 (1.1)
Italy	*	548 (2.2)	65 (1.4)	547 (2.1)	66 (1.3)	-1 (1.4)
Kazakhstan		536 (2.5)	63 (1.4)	533 (2.6)	65 (1.5)	-3 (1.0)
Latvia		558 (1.7)	62 (1.3)	556 (1.8)	64 (1.0)	-2 (0.8)
Lithuania		548 (2.6)	69 (1.7)	548 (2.4)	71 (1.5)	0 (1.0)
Macao SAR	*	546 (1.0)	66 (1.0)	546 (1.0)	66 (0.9)	0 (0.9)
Morocco		349 (4.3)	107 (2.3)	355 (4.4)	103 (1.8)	6 (1.8)
Netherlands	*	545 (1.7)	60 (1.2)	542 (1.8)	62 (1.3)	-3 (1.0)
New Zealand	*	523 (2.2)	91 (1.8)	522 (2.2)	90 (1.5)	-1 (0.8)
Northern Ireland		565 (2.2)	80 (1.3)	564 (2.3)	82 (1.3)	0 (0.8)
Norway (5)	*	559 (2.3)	65 (1.3)	558 (2.4)	65 (1.1)	-1 (1.2)
Oman	*	418 (3.3)	106 (1.7)	413 (3.4)	109 (1.6)	-5 (1.2)
Poland	*	565 (2.1)	72 (1.1)	565 (2.3)	75 (1.3)	0 (1.0)
Portugal	*	528 (2.3)	65 (1.4)	527 (2.2)	65 (1.4)	0 (0.8)
Qatar		442 (1.8)	110 (1.3)	440 (2.0)	110 (1.7)	-2 (0.7)
Russian Federation	*	581 (2.2)	66 (1.3)	581 (2.2)	69 (1.3)	0 (1.2)
Saudi Arabia		430 (4.2)	98 (2.4)	433 (4.0)	92 (2.2)	2 (1.1)
Singapore	*	576 (3.2)	80 (2.1)	576 (3.3)	85 (2.3)	0 (0.7)
Slovak Republic	*	535 (3.1)	81 (3.6)	534 (2.9)	80 (3.4)	-1 (1.0)
Slovenia	*	542 (2.0)	72 (1.1)	544 (2.1)	71 (1.3)	1 (1.3)
Spain	*	528 (1.7)	65 (1.4)	526 (1.6)	65 (1.2)	-1 (0.7)
Sweden	*	555 (2.4)	67 (1.2)	556 (2.5)	65 (1.1)	1 (0.9)
United Arab Emirates		450 (3.2)	111 (1.6)	447 (3.2)	112 (1.7)	-3 (0.7)
United States		549 (3.1)	78 (1.3)	548 (3.1)	80 (1.3)	-1 (1.2)
International Average		**524 (0.4)**	**77 (0.3)**	**524 (0.4)**	**77 (0.2)**	**0 (0.2)**
Benchmarking Participants						
Moscow City, Russian Fed.		612 (2.2)	62 (1.1)	615 (2.3)	65 (1.3)	3 (1.2)

Linking error is the standard error associated with the difference. Standard errors are shown in parentheses.

* Country contributed to item calibration.

Evaluating the Integration of ePIRLS with digitalPIRLS

The PIRLS 2021 reading achievement results for digitalPIRLS countries is based on the data collected from the 18 regular digitalPIRLS texts, which have a counterpart in paperPIRLS, as well as the five ePIRLS tasks. Two analyses were conducted to evaluate the integration of the five ePIRLS tasks into the achievement estimates of digitalPIRLS countries. First, item discrimination parameters were analyzed across digitalPIRLS and ePIRLS items. Then, the interim scale scores derived from only the digitalPIRLS items were compared to the final published scale scores for each digitalPIRLS country.

Item discrimination parameters were compared between digitalPIRLS items and ePIRLS items to examine the extent that ePIRLS items measure the same constructs as the regular digital items. Exhibit 11.28 shows summary statistics and plots of the estimated ePIRLS and digitalPIRLS item discrimination parameters. The discrimination parameters estimated for the ePIRLS items have about the same mean and range as the digitalPIRLS item discriminations, indicating that there is no functional difference between ePIRLS items and the regular digital items. If the ePIRLS items were measuring a different reading construct, their estimated discrimination parameters would be considerably smaller in value when aligned with the fixed digitalPIRLS scale.

Exhibit 11.28: **Item Discrimination Parameters for digitalPIRLS and ePIRLS Items**

	Count	Mean	Minimum	Maximum
digitalPIRLS	272	0.956	0.421	1.602
ePIRLS	91	0.946	0.341	1.953
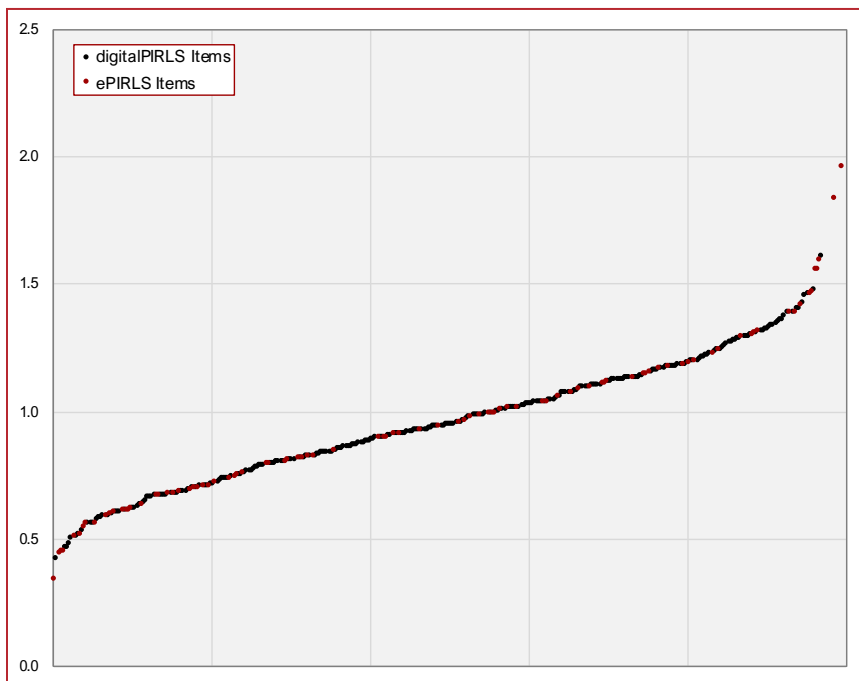
Exhibit 11.29 presents the achievement scale score estimates based on the two sets of items and the differences between them. Exhibit 11.30 shows a plot comparing the two sets of achievement scores.

Exhibit 11.29: PIRLS 2021 Average Reading Achievement for digitalPIRLS (Interim) and digitalPIRLS with ePIRLS

Country		digitalPIRLS Average Scale Score (not including ePIRLS)	digitalPIRLS with ePIRLS Average Scale Score	Difference	
Belgium (Flemish)	*	510 (2.4)	511 (2.3)	1 (1.2)	
Chinese Taipei	*	542 (2.1)	544 (2.2)	2 (1.0)	
Croatia		557 (3.0)	557 (2.5)	0 (1.9)	
Czech Republic	*	539 (2.8)	540 (2.3)	0 (1.2)	
Denmark	*	541 (2.3)	539 (2.2)	-2 (1.6)	
Finland	*	547 (2.7)	549 (2.4)	2 (1.2)	
Germany	*	528 (2.4)	524 (2.1)	-4 (1.3)	▼
Hungary		539 (3.4)	539 (3.4)	0 (1.3)	
Israel		511 (2.5)	510 (2.2)	-1 (1.2)	
Italy	*	536 (2.1)	537 (2.2)	1 (1.0)	
Kazakhstan		503 (3.0)	504 (2.7)	0 (1.1)	
Lithuania		555 (2.5)	552 (2.3)	-3 (1.6)	
Malta		514 (2.8)	515 (2.7)	1 (1.3)	
New Zealand	*	519 (2.8)	521 (2.3)	2 (1.4)	
Norway (5)	*	538 (2.1)	539 (2.0)	1 (1.0)	
Portugal	*	518 (2.4)	520 (2.3)	2 (1.6)	
Qatar		483 (4.0)	485 (3.7)	1 (1.4)	
Russian Federation	*	567 (3.7)	567 (3.6)	0 (1.2)	
Saudi Arabia		446 (3.8)	449 (3.6)	2 (1.8)	
Singapore	*	589 (3.1)	587 (3.1)	-2 (0.9)	▼
Slovak Republic	*	529 (2.8)	529 (2.7)	0 (1.5)	
Slovenia	*	523 (2.2)	520 (1.9)	-4 (1.1)	▼
Spain	*	518 (2.3)	521 (2.2)	3 (1.2)	▲
Sweden	*	544 (2.6)	544 (2.1)	0 (1.3)	
United Arab Emirates		479 (2.0)	483 (1.8)	5 (0.7)	▲
International Average		**527 (0.6)**	**527 (0.5)**	**0 (0.3)**	

Exhibit 11.29: **PIRLS 2021 Average Reading Achievement for digitalPIRLS (Interim) and digitalPIRLS with ePIRLS (Continued)**

Country	digitalPIRLS Average Scale Score (not including ePIRLS)	digitalPIRLS with ePIRLS Average Scale Score	Difference	
Benchmarking Participants				
Alberta, Canada	538 (3.5)	539 (3.6)	1 (1.7)	
British Columbia, Canada	533 (3.4)	535 (3.5)	2 (1.5)	
Newfoundland & Labrador, Can.	524 (3.5)	523 (3.2)	0 (1.6)	
Quebec, Canada	555 (3.1)	551 (2.7)	-4 (1.4)	▼
Moscow City, Russian Fed.	598 (2.2)	598 (2.1)	0 (1.2)	
Abu Dhabi, UAE	435 (4.1)	440 (3.5)	4 (1.4)	▲
Dubai, UAE	549 (1.7)	552 (1.5)	3 (1.2)	▲

( ) Standard errors appear in parentheses. Because of rounding some results may appear inconsistent.

* Country contributed to item calibration.

▲ indicates the published scores are significantly higher than the interim digitalPIRLS scores (α = 0.05)

▼ indicates the published scores are significantly lower than the interim digitalPIRLS scores (α = 0.05)
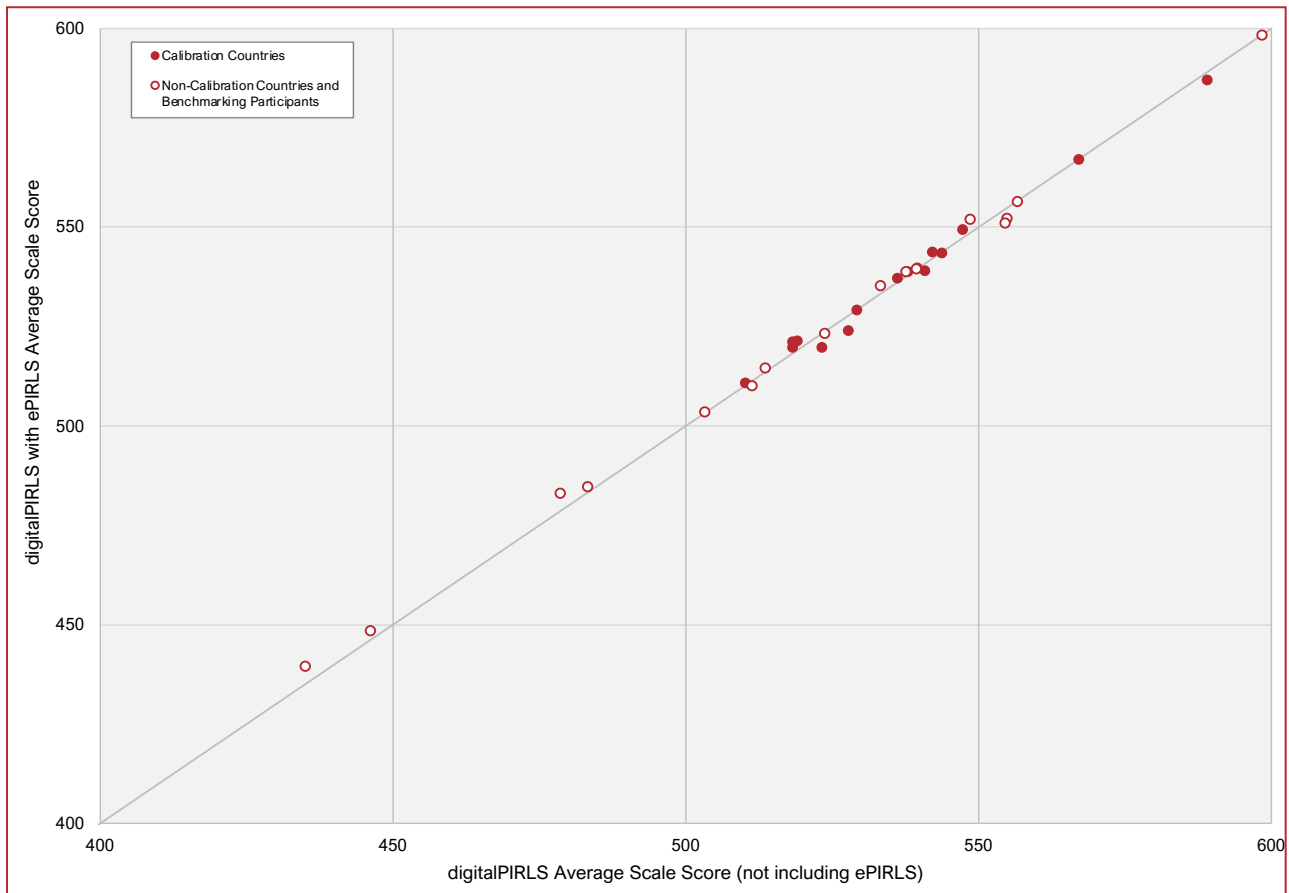
There were very small differences between the two sets of results, most being two points or less. The high degree of agreement between the two sets of achievement results provides evidence that the scaling approach for integrating ePIRLS and digitalPIRLS is valid and provides an overall measure of reading achievement that is consistent with the PIRLS Reading Framework and comparable with previous PIRLS assessments.

Conclusion

Scaling the PIRLS 2021 achievement data was successful in estimating valid and comparable plausible values from its paperPIRLS and digitalPIRLS assessments. The psychometric methods implemented and described in this chapter relied on past methods and experience for scaling the paperPIRLS data. The use of multiple-group IRT models enabled PIRLS to find international item parameters that maximize fit across all countries. Scaling the digitalPIRLS data required careful

consideration of any potential mode differences as well as the complex data collection conditions in PIRLS 2021 due to the COVID-19 pandemic.

The major outcome was the successful linking of all PIRLS 2021 assessment data to the PIRLS trend scale such that results from the paper-based and computer-based 2021 assessments can be compared directly without further need for adjustments. They also can be compared reliably with past PIRLS assessments.

References

Beaton, A. E. (1969). Criterion scaling of questionnaire items. *Socio-Economic Planning Sciences, 2*, 355–362.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Dorans, N. J., & Puhan, G. (2017). Contributions to score linking theory and practice. In R. Bennett, & M. von Davier (Eds.), *Advancing human assessment. Methodology of educational measurement and assessment*. Springer, Cham. https://doi.org/10.1007/978-3-319-58689-2_4

Foy, P., & Yin, L. (2017). Scaling the PIRLS 2016 achievement data. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016* (pp. 12.1-12.38). Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/publications/pirls/2016-methods/chapter-12.html

Glas, C. A. W., & Pimentel, J. (2006). *Modeling nonignorable missing data processes in item calibration*. (LSAC Research Report Series; No. 04-07). Newton, PA: Law School Admission Council.

Haberman, S. J. (2015). Pseudo-equivalent groups and linking. *Journal of Educational and Behavioral Statistics, 40*(3), 254–273. https://doi.org/10.3102/1076998615574772

Kim, S. (2006). A Comparative Study of IRT fixed parameter calibration methods. *Journal of Educational Measurement, 43*(4), 355–381. http://www.jstor.org/stable/20461835

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York: Springer Science+Business Media.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: J. Wiley & Sons.

Martin, M. O., von Davier, M., Foy, P., & Mullis, I. V. S. (2019). PIRLS 2021 assessment design. In I. V. S. Mullis & M.O. Martin (Eds.), *PIRLS 2021 assessment frameworks*. Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/pirls2021/frameworks/

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*, 359-381.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika, 56*, 177-196.

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.

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

CHAPTER 11: IMPLEMENTING THE SCALING METHODOLOGY
METHODS AND PROCEDURES: PIRLS 2021 TECHNICAL REPORT   11.49

Moustaki, I., & Knott, M. (2000). Weighted for item non-response in attitude scales by using latent variable models with covariates. *Journal of the Royal Statistical Society, 163*(3), 445–459.

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*. Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/pirls2021/frameworks/

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/index.html

Rogers, A., Tang, C., Lin, M.-J., & Kandathil, M. (2006). DGROUP [computer software]. Princeton, NJ: Educational Testing Service.

Rose, N., von Davier, M., & Xu, X. (2010). *Modeling nonignorable missing data with item response theory (IRT).* (ETS Research Report No. RR-10-11). Princeton, NJ: Educational Testing Service.

Rose, N., von Davier, M., & Nagengast, B. (2017). Modeling omitted and not-reached items in IRT models. *Psychometrika, 82*(3), 795-819.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592. https://www.jstor.org/stable/pdf/2335739.pdf

Sheehan, K. M. (1985). M-Group: Estimation of group effects in multivariate models [computer software, Version 3.2]. Princeton, NJ: Educational Testing Service.

Thomas, N. (1993). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics, 2*, 309 322.

Vale, C. D. (1986). Linking item parameters onto a common scale. *Applied Psychological Measurement, 10*(4), 333–344. https://doi.org/10.1177/014662168601000402

von Davier, M., & Bezirhan, U. (2022). A robust method for detecting item misfit in large-scale assessments. *Educational and Psychological Measurement*. https://doi.org/10.1177/00131644221105819

von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? In M. von Davier & D. Hastedt (Eds.), *IERI Monograph Series: Issues and Methodologies in Large Scale Assessments* (Vol. 2, pp. 9-36). Retrieved from https://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_01.pdf

von Davier, M., & Sinharay, S. (2013). 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, FL: CRC Press.

von Davier, M., Sinharay, S., Oranje, A., & Beaton, A. (2006). Statistical procedures used in the National Assessment of Educational Progress (NAEP): Recent developments and future directions. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. 26: Psychometrics). Amsterdam, Netherlands: Elsevier.

von Davier, M., & von Davier, A. A. (2007). A unified approach to IRT scale linking and scale transformations. Methodology: *European Journal of Research Methods for the Behavioral and Social Sciences, 3*(3), 115-124.

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

CHAPTER 11: IMPLEMENTING THE SCALING METHODOLOGY
METHODS AND PROCEDURES: PIRLS 2021 TECHNICAL REPORT   11.50

Appendix 11A: Item Parameters from the PIRLS 2021 Paper Concurrent Calibration

Paper item parameters estimated from the PIRLS 2021 concurrent calibration. Used to estimate student proficiency for paperPIRLS countries and bridge data from digitalPIRLS countries.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
				Items Released in 2016				
R11F01M	0.022	—	1.253 (0.078)	-0.602 (0.055)	0.107 (0.014)			
R11F02M	0.015	—	0.704 (0.047)	-0.846 (0.089)	0.209 (0.026)			
R11F03M	0.014	—	1.038 (0.065)	-0.607 (0.077)	0.153 (0.019)			
R11F04M	0.011	—	1.353 (0.050)	-0.756 (0.058)	0.170 (0.018)			
R11F05M	0.020	—	0.951 (0.062)	-0.276 (0.119)	0.185 (0.026)			
R11F06C	0.012	—	0.753 (0.054)	-0.082 (0.081)				
R11F07C	0.023	—	0.459 (0.034)	0.361 (0.078)		-1.110 (0.147)	1.110 (0.147)	
R11F08C	0.032	—	1.029 (0.037)	-0.334 (0.068)				
R11F09C	0.015	—	0.990 (0.059)	-0.604 (0.074)		0.100 (0.039)	-0.100 (0.039)	
R11F10C	0.011	—	0.790 (0.067)	-1.508 (0.170)				
R11F11M	0.011	—	0.631 (0.061)	0.110 (0.108)	0.175 (0.021)			
R11F12C	0.017	—	0.635 (0.030)	0.572 (0.063)		-0.402 (0.095)	0.402 (0.095)	
R11F13M	0.014	—	1.008 (0.064)	-0.317 (0.091)	0.193 (0.034)			
R11L01M	0.026	—	0.612 (0.049)	-1.957 (0.152)	0.177 (0.020)			
R11L02M	0.018	—	0.653 (0.050)	0.496 (0.089)	0.180 (0.026)			
R11L03C	0.018	—	0.638 (0.041)	-0.462 (0.082)				
R11L04C	0.038	—	0.717 (0.038)	0.347 (0.065)		1.957 (0.104)	-1.162 (0.105)	-0.795 (0.062)
R11L05M	0.022	—	1.062 (0.072)	0.463 (0.087)	0.184 (0.019)			
R11L06C	0.015	—	0.622 (0.039)	0.302 (0.092)				
R11L07M	0.021	—	0.702 (0.068)	0.524 (0.095)	0.151 (0.015)			
R11L08C	0.021	—	0.810 (0.040)	0.492 (0.066)		0.856 (0.053)	-0.856 (0.053)	
R11L09M	0.014	—	0.988 (0.063)	-0.920 (0.112)	0.203 (0.026)			
R11L10C	0.024	—	0.685 (0.038)	0.636 (0.081)		0.397 (0.073)	-0.397 (0.073)	
R11L11M	0.018	—	0.901 (0.050)	-0.306 (0.100)	0.181 (0.022)			
R11L12C	0.016	—	0.706 (0.051)	0.405 (0.091)		1.062 (0.090)	-1.062 (0.090)	
R41H01M	0.013	—	0.817 (0.075)	-1.108 (0.329)	0.437 (0.159)			
R41H02M	0.013	—	1.026 (0.046)	-0.748 (0.074)	0.185 (0.033)			

Paper item parameters estimated from the PIRLS 2021 concurrent calibration. Used to estimate student proficiency for paperPIRLS countries and bridge data from digitalPIRLS countries.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
R41H03C	0.016	–	1.143 (0.066)	0.194 (0.053)				
R41H04C	0.017	–	0.645 (0.056)	1.626 (0.159)				
R41H05M	0.009	–	0.876 (0.090)	0.131 (0.104)	0.268 (0.028)			
R41H06C	0.012	–	0.687 (0.037)	-0.390 (0.083)		0.095 (0.051)	-0.095 (0.051)	
R41H07M	0.021	–	0.801 (0.055)	0.799 (0.097)	0.122 (0.011)			
R41H08C	0.019	–	0.670 (0.039)	1.472 (0.106)				
R41H09M	0.017	–	0.665 (0.059)	0.175 (0.091)	0.161 (0.027)			
R41H10M	0.018	–	1.218 (0.047)	-0.240 (0.066)	0.217 (0.027)			
R41H11M	0.010	–	1.253 (0.049)	-0.677 (0.089)	0.253 (0.044)			
R41H12M	0.012	–	1.216 (0.051)	-0.102 (0.070)	0.318 (0.022)			
R41H13C	0.018	–	0.510 (0.030)	0.741 (0.076)		0.158 (0.127)	0.166 (0.116)	-0.323 (0.075)
R41H14C	0.029	–	0.879 (0.050)	0.157 (0.071)				
R41H15C	0.020	–	1.205 (0.041)	-0.455 (0.059)				
R41H16C	0.017	–	0.916 (0.048)	0.599 (0.085)				
R41T01M	0.008	–	1.066 (0.044)	-1.087 (0.084)	0.222 (0.030)			
R41T02C	0.018	–	0.688 (0.044)	-0.629 (0.085)		0.386 (0.038)	-0.386 (0.038)	
R41T03C	0.014	–	0.955 (0.033)	0.020 (0.054)		0.081 (0.022)	-0.081 (0.022)	
R41T04C	0.007	–	1.090 (0.045)	-0.124 (0.070)				
R41T05M	0.027	–	0.740 (0.072)	0.351 (0.131)	0.244 (0.050)			
R41T06C	0.014	–	1.332 (0.056)	-0.760 (0.066)				
R41T07C	0.020	–	0.745 (0.032)	0.633 (0.057)		-0.221 (0.045)	0.221 (0.045)	
R41T08C	0.010	–	1.108 (0.039)	-0.110 (0.072)				
R41T09M	0.023	–	1.263 (0.068)	0.420 (0.068)	0.146 (0.018)			
R41T10C	0.010	–	1.227 (0.047)	-0.370 (0.068)				
R41T11C	0.020	–	0.721 (0.033)	0.504 (0.061)		-0.399 (0.068)	0.315 (0.049)	0.084 (0.059)
R41T12M	0.009	–	0.908 (0.054)	-0.361 (0.148)	0.260 (0.083)			
R41T13M	0.011	–	1.084 (0.054)	0.402 (0.069)	0.110 (0.009)			
R41T14C	0.017	–	0.445 (0.032)	0.012 (0.119)				
R41T15M	0.013	–	0.846 (0.049)	0.231 (0.083)	0.199 (0.027)			
R41T16M	0.023	–	1.119 (0.067)	-0.207 (0.069)	0.204 (0.027)			
Items Common in 2016 and 2021								
RP21K01	0.018	0.025	0.464 (0.035)	-0.818 (0.101)		0.240 (0.051)	-0.240 (0.051)	
RP21K02	0.026	0.026	0.765 (0.044)	-0.707 (0.091)				
RP21K03	0.024	0.011	0.875 (0.059)	-0.062 (0.089)	0.153 (0.024)			
RP21K05	0.018	0.026	1.029 (0.049)	0.175 (0.076)				

Paper item parameters estimated from the PIRLS 2021 concurrent calibration. Used to estimate student proficiency for paperPIRLS countries and bridge data from digitalPIRLS countries.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP21K06	0.010	0.025	1.378 (0.085)	-0.039 (0.077)	0.261 (0.021)			
RP21K07	0.029	0.033	0.777 (0.043)	0.101 (0.066)		0.229 (0.032)	-0.229 (0.032)	
RP21K08	0.026	0.013	0.955 (0.066)	0.304 (0.076)	0.150 (0.030)			
RP21K09	0.018	0.024	1.232 (0.061)	-0.092 (0.076)	0.198 (0.021)			
RP21K10	0.012	0.019	0.681 (0.035)	0.652 (0.049)		-0.290 (0.089)	0.290 (0.089)	
RP21K11	0.017	0.015	1.006 (0.087)	0.018 (0.074)	0.164 (0.027)			
RP21K12	0.023	0.022	0.585 (0.030)	-0.246 (0.071)		0.861 (0.066)	-0.270 (0.052)	-0.591 (0.046)
RP21Y01	0.020	0.014	1.077 (0.058)	0.156 (0.076)	0.215 (0.024)			
RP21Y02	0.019	0.013	1.563 (0.062)	-0.265 (0.062)	0.235 (0.024)			
RP21Y03	0.039	0.030	0.629 (0.050)	0.481 (0.096)				
RP21Y04	0.021	0.017	1.178 (0.072)	0.021 (0.062)	0.165 (0.021)			
RP21Y05	0.014	0.019	1.582 (0.076)	0.075 (0.075)	0.208 (0.017)			
RP21Y06	0.023	0.016	1.508 (0.063)	0.042 (0.049)	0.183 (0.017)			
RP21Y07	0.030	0.018	0.847 (0.062)	-0.901 (0.074)	0.129 (0.019)			
RP21Y08	0.031	0.012	1.470 (0.051)	-0.177 (0.062)	0.225 (0.029)			
RP21Y09	0.020	0.015	1.006 (0.044)	-0.582 (0.065)		0.221 (0.024)	-0.221 (0.024)	
RP21Y10	0.027	0.024	0.661 (0.031)	0.537 (0.074)				
RP21Y11	0.014	0.022	1.283 (0.071)	-0.024 (0.055)	0.269 (0.019)			
RP21Y12	0.018	0.012	0.696 (0.029)	0.062 (0.058)		-1.142 (0.124)	1.142 (0.124)	
RP21Y13	0.027	0.014	0.757 (0.028)	0.283 (0.059)		0.980 (0.070)	-0.476 (0.081)	-0.504 (0.048)
RP21Y14	0.021	0.020	0.547 (0.027)	0.284 (0.067)		-0.561 (0.095)	0.561 (0.095)	
RP31M01	0.013	0.012	1.417 (0.077)	-0.941 (0.080)	0.167 (0.026)			
RP31M02	0.011	0.008	1.221 (0.055)	-1.072 (0.082)				
RP31M03	0.013	0.014	1.189 (0.057)	-0.146 (0.066)	0.200 (0.031)			
RP31M04	0.021	0.019	0.570 (0.051)	0.247 (0.109)				
RP31M05	0.024	0.029	1.266 (0.109)	0.058 (0.065)	0.384 (0.027)			
RP31M06	0.013	0.027	1.077 (0.071)	0.193 (0.079)	0.252 (0.020)			
RP31M07	0.030	0.010	1.446 (0.080)	-0.372 (0.064)	0.184 (0.019)			
RP31M08	0.021	0.015	1.342 (0.046)	-0.588 (0.063)	0.185 (0.031)			
RP31M09	0.038	0.032	0.678 (0.035)	-0.262 (0.063)		1.434 (0.089)	-1.434 (0.089)	
RP31M10	0.042	0.043	0.564 (0.053)	0.254 (0.132)				
RP31M11	0.033	0.015	0.855 (0.065)	-0.807 (0.120)	0.203 (0.038)			
RP31M12	0.021	0.021	1.136 (0.054)	-0.022 (0.066)	0.125 (0.016)			
RP31M13	0.010	0.014	1.885 (0.058)	-0.784 (0.074)	0.189 (0.020)			
RP31M14	0.024	0.023	1.992 (0.069)	-0.334 (0.060)	0.157 (0.013)			

Paper item parameters estimated from the PIRLS 2021 concurrent calibration. Used to estimate student proficiency for paperPIRLS countries and bridge data from digitalPIRLS countries.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP31M15	0.016	0.023	1.233 (0.083)	-0.138 (0.077)	0.193 (0.016)			
RP31M16	0.022	0.035	1.195 (0.046)	-0.066 (0.060)				
RP31M17	0.021	0.020	0.577 (0.030)	-0.157 (0.058)		0.363 (0.058)	0.101 (0.036)	-0.464 (0.043)
RP31W01	0.044	0.044	0.789 (0.038)	-0.547 (0.061)		0.241 (0.038)	-0.241 (0.038)	
RP31W02	0.019	0.032	0.773 (0.032)	0.231 (0.055)		-0.092 (0.038)	0.092 (0.038)	
RP31W03	0.026	0.008	1.309 (0.101)	-0.047 (0.081)	0.151 (0.025)			
RP31W04	0.028	0.018	0.856 (0.050)	-0.719 (0.086)				
RP31W05	0.029	0.007	1.160 (0.069)	0.385 (0.073)	0.243 (0.021)			
RP31W06	0.026	0.009	0.762 (0.041)	-0.998 (0.133)	0.169 (0.027)			
RP31W07	0.017	0.017	0.925 (0.037)	0.487 (0.076)		0.018 (0.031)	0.116 (0.027)	-0.134 (0.029)
RP31W08	0.010	0.016	1.265 (0.060)	-0.214 (0.074)	0.240 (0.026)			
RP31W09	0.025	0.023	0.840 (0.103)	0.381 (0.086)	0.132 (0.023)			
RP31W10	0.014	0.016	1.259 (0.070)	0.285 (0.058)	0.162 (0.015)			
RP31W11	0.014	0.019	1.367 (0.048)	0.452 (0.068)				
RP31W12	0.021	0.013	1.421 (0.062)	0.621 (0.074)	0.196 (0.012)			
RP31W13	0.023	0.016	0.781 (0.035)	0.689 (0.069)				
RP41B01	0.009	0.017	0.747 (0.056)	-2.303 (0.184)				
RP41B02	0.020	0.013	1.078 (0.060)	-1.951 (0.109)	0.157 (0.024)			
RP41B03	0.022	0.019	0.746 (0.064)	-2.343 (0.192)	0.167 (0.030)			
RP41B04	0.020	0.015	0.802 (0.041)	-1.041 (0.082)				
RP41B05	0.020	0.009	0.991 (0.045)	-1.254 (0.065)	0.164 (0.027)			
RP41B06	0.018	0.025	0.546 (0.048)	-1.839 (0.219)	0.174 (0.033)			
RP41B07	0.026	0.019	0.796 (0.051)	-1.050 (0.112)				
RP41B08	0.017	0.009	0.843 (0.051)	-1.409 (0.116)				
RP41B09	0.020	0.010	1.012 (0.045)	-0.948 (0.087)				
RP41B10	0.008	0.019	0.649 (0.043)	-0.432 (0.129)	0.183 (0.029)			
RP41B11	0.020	0.006	1.022 (0.059)	-1.125 (0.123)	0.207 (0.040)			
RP41B12	0.014	0.012	0.845 (0.062)	-1.917 (0.123)	0.182 (0.038)			
RP41B13	0.028	0.040	0.502 (0.033)	-1.130 (0.115)		0.843 (0.069)	-0.843 (0.069)	
RP41B14	0.022	0.025	1.033 (0.055)	-0.689 (0.074)	0.134 (0.017)			
RP41B15	0.028	0.027	0.351 (0.027)	0.310 (0.116)				
RP41B16	0.024	0.026	0.428 (0.025)	-0.190 (0.077)		0.147 (0.052)	-0.147 (0.052)	
RP41B17	0.026	0.019	0.717 (0.036)	-0.432 (0.066)		0.900 (0.061)	-0.900 (0.061)	
RP41E01	0.020	0.016	0.729 (0.068)	-3.153 (0.159)				
RP41E02	0.030	0.019	0.801 (0.057)	-1.701 (0.155)				

Paper item parameters estimated from the PIRLS 2021 concurrent calibration. Used to estimate student proficiency for paperPIRLS countries and bridge data from digitalPIRLS countries.

Item	RMSD 2016	2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP41E03	0.013	0.031	1.284 (0.061)	-0.691 (0.101)	0.251 (0.036)			
RP41E04	0.008	0.016	1.058 (0.053)	-0.466 (0.086)	0.206 (0.025)			
RP41E05	0.014	0.010	1.266 (0.051)	-1.427 (0.088)	0.178 (0.037)			
RP41E06	0.012	0.026	1.075 (0.071)	-2.186 (0.160)	0.274 (0.068)			
RP41E07	0.027	0.021	0.512 (0.023)	-0.839 (0.091)		0.104 (0.064)	-0.104 (0.064)	
RP41E08	0.021	0.006	0.968 (0.061)	-0.909 (0.123)	0.178 (0.032)			
RP41E09	0.023	0.020	0.530 (0.103)	0.396 (0.301)	0.332 (0.071)			
RP41E10	0.030	0.022	0.828 (0.062)	-3.015 (0.155)				
RP41E11	0.026	0.007	1.363 (0.060)	-0.750 (0.080)	0.235 (0.026)			
RP41E12	0.023	0.012	1.128 (0.057)	-2.377 (0.118)				
RP41E13	0.026	0.017	0.516 (0.037)	-0.292 (0.097)				
RP41E14	0.026	0.026	0.458 (0.027)	0.000 (0.089)		1.032 (0.075)	-1.032 (0.075)	
RP41E15	0.015	0.015	0.770 (0.033)	-2.368 (0.096)				
RP41E16	0.030	0.015	0.782 (0.033)	-0.906 (0.077)				
RP41E17	0.014	0.012	1.117 (0.065)	-0.502 (0.070)	0.146 (0.019)			
RP41I01	0.022	0.010	0.940 (0.050)	-1.355 (0.114)				
RP41I02	0.030	0.016	0.954 (0.072)	0.494 (0.098)	0.149 (0.017)			
RP41I03	0.023	0.023	0.538 (0.022)	0.183 (0.067)		0.413 (0.030)	-0.413 (0.030)	
RP41I04	0.024	0.025	0.702 (0.025)	0.346 (0.063)		0.476 (0.041)	-0.476 (0.041)	
RP41I05	0.030	0.020	1.067 (0.061)	0.041 (0.083)	0.195 (0.035)			
RP41I06	0.032	0.018	1.191 (0.071)	0.451 (0.098)	0.227 (0.027)			
RP41I07	0.016	0.011	0.961 (0.033)	0.424 (0.045)		0.040 (0.034)	-0.040 (0.034)	
RP41I08	0.037	0.011	1.568 (0.072)	-0.574 (0.073)	0.226 (0.035)			
RP41I09	0.034	0.007	1.137 (0.057)	-0.096 (0.068)				
RP41I10	0.013	0.018	0.841 (0.076)	-0.004 (0.110)	0.168 (0.030)			
RP41I11	0.028	0.018	0.864 (0.038)	0.393 (0.058)		0.443 (0.036)	-0.443 (0.036)	
RP41I12	0.019	0.016	1.085 (0.048)	0.190 (0.053)	0.130 (0.017)			
RP41I13	0.019	0.024	0.864 (0.040)	0.177 (0.067)				
RP41I14	0.014	0.011	0.786 (0.049)	0.845 (0.070)				
RP41I15	0.023	0.008	0.858 (0.043)	0.346 (0.059)				
RP41O01	0.015	0.012	0.970 (0.044)	-0.976 (0.095)	0.213 (0.031)			
RP41O02	0.010	0.010	0.931 (0.057)	-0.790 (0.063)				
RP41O03	0.024	0.027	0.908 (0.056)	0.876 (0.089)				
RP41O04	0.020	0.026	0.542 (0.046)	1.154 (0.101)		-0.049 (0.074)	0.049 (0.074)	
RP41O05	0.025	0.030	0.603 (0.032)	0.192 (0.069)		-0.213 (0.074)	0.213 (0.074)	

Paper item parameters estimated from the PIRLS 2021 concurrent calibration. Used to estimate student proficiency for paperPIRLS countries and bridge data from digitalPIRLS countries.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP41O06	0.011	0.019	1.404 (0.058)	-0.277 (0.067)	0.262 (0.022)			
RP41O07	0.033	0.032	0.609 (0.032)	-1.073 (0.076)		0.217 (0.039)	-0.217 (0.039)	
RP41O08	0.011	0.027	0.771 (0.049)	-0.106 (0.087)				
RP41O09	0.020	0.041	1.128 (0.045)	-0.149 (0.059)				
RP41O10	0.026	0.017	0.792 (0.032)	0.321 (0.062)		-0.023 (0.040)	0.023 (0.040)	
RP41O11	0.026	0.014	1.356 (0.057)	0.118 (0.059)	0.259 (0.026)			
RP41O12	0.030	0.012	1.320 (0.061)	-0.308 (0.054)	0.137 (0.018)			
RP41O13	0.020	0.020	0.536 (0.040)	0.347 (0.064)		-1.189 (0.225)	0.920 (0.161)	0.269 (0.097)
Items Introduced in 2021								
* RP31P01	—	0.010	0.835 (0.116)	-0.780 (0.211)	0.226 (0.051)			
* RP31P02	—	0.005	1.029 (0.051)	-1.517 (0.182)	0.195 (0.015)			
* RP31P03	—	0.020	0.788 (0.074)	-0.925 (0.127)		0.507 (0.079)	-0.507 (0.079)	
* RP31P04	—	0.009	0.955 (0.042)	-2.082 (0.177)				
* RP31P05	—	0.019	1.228 (0.068)	-1.215 (0.159)	0.193 (0.026)			
* RP31P06	—	0.019	0.651 (0.098)	-2.502 (0.307)				
* RP31P07	—	0.029	0.733 (0.070)	-0.703 (0.097)				
* RP31P08	—	0.020	1.064 (0.099)	-1.378 (0.211)	0.196 (0.028)			
* RP31P09	—	0.020	1.236 (0.061)	-1.420 (0.121)	0.216 (0.024)			
* RP31P10	—	0.010	1.101 (0.042)	-1.432 (0.114)	0.192 (0.015)			
* RP31P11	—	0.019	0.749 (0.065)	-1.064 (0.135)		-0.263 (0.143)	0.263 (0.143)	
* RP31P12	—	0.030	0.566 (0.114)	-0.466 (0.247)				
* RP31P13	—	0.016	1.116 (0.053)	-1.106 (0.111)				
* RP31P14	—	0.011	1.023 (0.047)	-0.674 (0.080)				
* RP31U01	—	0.008	0.741 (0.061)	-2.466 (0.183)				
* RP31U02	—	0.011	0.912 (0.063)	-2.296 (0.129)	0.248 (0.016)			
* RP31U03	—	0.008	0.973 (0.060)	-1.510 (0.120)	0.195 (0.019)			
* RP31U04	—	0.020	0.622 (0.100)	-1.504 (0.297)				
* RP31U05	—	0.020	0.935 (0.068)	-1.467 (0.141)	0.199 (0.031)			
* RP31U06	—	0.016	1.043 (0.068)	-1.446 (0.121)	0.181 (0.020)			
* RP31U07	—	0.014	1.060 (0.070)	-0.438 (0.092)	0.172 (0.019)			
* RP31U08	—	0.013	0.986 (0.097)	-0.608 (0.132)	0.183 (0.019)			
* RP31U09	—	0.009	1.074 (0.045)	-1.491 (0.082)	0.214 (0.023)			
* RP31U10	—	0.020	0.724 (0.090)	-0.554 (0.163)				
* RP31U11	—	0.020	0.722 (0.051)	-1.521 (0.081)		0.204 (0.120)	-0.204 (0.120)	
* RP31U12	—	0.019	1.070 (0.063)	-1.421 (0.132)				

Paper item parameters estimated from the PIRLS 2021 concurrent calibration. Used to estimate student proficiency for paperPIRLS countries and bridge data from digitalPIRLS countries.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
* RP31U13	—	0.022	1.172 (0.043)	-1.588 (0.071)	0.169 (0.013)			
* RP31U14	—	0.043	0.522 (0.097)	0.265 (0.180)				
* RP41H01	—	0.010	0.890 (0.043)	-2.422 (0.149)				
* RP41H02	—	0.017	1.082 (0.048)	-1.507 (0.085)	0.195 (0.017)			
* RP41H03	—	0.020	0.671 (0.087)	-0.183 (0.174)	0.185 (0.023)			
* RP41H04	—	0.023	0.747 (0.084)	-1.852 (0.155)	0.212 (0.018)			
* RP41H05	—	0.023	1.007 (0.090)	-2.191 (0.288)	0.260 (0.023)			
* RP41H06	—	0.017	0.887 (0.116)	-2.103 (0.243)				
* RP41H07	—	0.014	1.141 (0.052)	-0.997 (0.093)	0.197 (0.022)			
* RP41H08	—	0.010	1.051 (0.062)	-0.849 (0.123)	0.163 (0.014)			
* RP41H09	—	0.020	1.048 (0.055)	-1.478 (0.100)	0.199 (0.020)			
* RP41H10	—	0.008	0.967 (0.063)	-0.912 (0.091)	0.203 (0.027)			
* RP41H11	—	0.010	0.694 (0.072)	-1.649 (0.194)				
* RP41H12	—	0.022	0.971 (0.063)	-1.079 (0.102)				
* RP41H13	—	0.008	1.133 (0.052)	-1.195 (0.094)	0.175 (0.025)			
* RP41H14	—	0.030	0.866 (0.143)	-1.768 (0.135)				
* RP41H15	—	0.012	0.922 (0.056)	-0.540 (0.131)	0.170 (0.021)			
* RP41H16	—	0.010	1.092 (0.046)	-0.888 (0.106)	0.187 (0.018)			
* RP41M01	—	0.021	0.636 (0.152)	-2.525 (0.542)	0.289 (0.040)			
* RP41M02	—	0.014	1.088 (0.051)	-0.250 (0.133)	0.170 (0.025)			
* RP41M03	—	0.019	0.670 (0.057)	-2.136 (0.172)				
* RP41M04	—	0.014	0.858 (0.109)	-1.086 (0.104)				
* RP41M05	—	0.014	0.821 (0.068)	-2.242 (0.190)	0.225 (0.019)			
* RP41M06	—	0.009	0.938 (0.153)	-1.901 (0.596)				
* RP41M07	—	0.007	0.958 (0.064)	-1.133 (0.141)	0.228 (0.027)			
* RP41M08	—	0.023	1.067 (0.066)	-2.160 (0.173)				
* RP41M09	—	0.007	0.963 (0.060)	-1.323 (0.099)	0.208 (0.034)			
* RP41M10	—	0.009	0.712 (0.103)	-1.789 (0.253)		0.183 (0.085)	-0.183 (0.085)	
* RP41M11	—	0.009	1.080 (0.054)	-0.524 (0.105)	0.180 (0.035)			
* RP41M12	—	0.018	0.804 (0.182)	-1.782 (0.298)				
* RP41M13	—	0.024	1.047 (0.082)	-2.184 (0.207)	0.203 (0.034)			
* RP41M14	—	0.019	0.647 (0.039)	-0.641 (0.124)		0.667 (0.146)	0.067 (0.083)	-0.735 (0.094)
* RP41M15	—	0.010	1.034 (0.082)	-0.831 (0.128)	0.208 (0.031)			
* RP41M16	—	0.009	0.893 (0.093)	-1.542 (0.181)				
* RP41M17	—	0.013	1.132 (0.096)	-0.491 (0.121)				

Paper item parameters estimated from the PIRLS 2021 concurrent calibration. Used to estimate student proficiency for paperPIRLS countries and bridge data from digitalPIRLS countries.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
* RP41M18	—	0.023	1.137 (0.117)	-0.789 (0.114)	0.148 (0.019)			
RP51C01	—	0.022	0.460 (0.106)	0.480 (0.135)		0.096 (0.088)	-0.096 (0.088)	
RP51C02	—	0.011	0.980 (0.058)	-0.786 (0.089)	0.167 (0.021)			
RP51C03	—	0.015	0.954 (0.055)	0.538 (0.106)				
RP51C04	—	0.009	0.978 (0.062)	-1.158 (0.123)	0.199 (0.027)			
RP51C05	—	0.021	0.983 (0.047)	-0.655 (0.062)		0.372 (0.046)	-0.372 (0.046)	
RP51C06	—	0.023	0.595 (0.064)	0.409 (0.141)		-0.065 (0.161)	0.065 (0.161)	
RP51C07	—	0.011	0.696 (0.088)	-0.474 (0.145)				
RP51C08	—	0.013	0.877 (0.061)	-0.235 (0.106)	0.177 (0.026)			
RP51C09	—	0.023	1.232 (0.062)	-1.018 (0.097)	0.162 (0.016)			
RP51C10	—	0.021	0.837 (0.050)	-0.225 (0.073)		0.659 (0.077)	-0.659 (0.077)	
RP51C11	—	0.018	0.943 (0.063)	-0.285 (0.097)	0.158 (0.016)			
RP51C12	—	0.014	1.186 (0.056)	-0.651 (0.075)	0.190 (0.024)			
RP51C13	—	0.025	0.867 (0.053)	-0.011 (0.089)		0.361 (0.056)	-0.361 (0.056)	
RP51C14	—	0.027	0.934 (0.067)	0.852 (0.080)		0.176 (0.072)	-0.176 (0.072)	
RP51C15	—	0.020	0.742 (0.096)	1.213 (0.154)				
RP51D01	—	0.006	0.934 (0.078)	-1.040 (0.109)				
RP51D02	—	0.022	0.751 (0.080)	0.105 (0.123)	0.194 (0.031)			
RP51D03	—	0.013	0.884 (0.057)	-0.182 (0.105)				
RP51D04	—	0.012	1.050 (0.103)	-0.343 (0.131)	0.150 (0.016)			
RP51D05	—	0.019	0.871 (0.062)	-0.351 (0.107)	0.177 (0.018)			
RP51D06	—	0.031	0.700 (0.050)	-0.098 (0.083)		0.636 (0.066)	-0.636 (0.066)	
RP51D07	—	0.010	1.003 (0.081)	0.368 (0.089)				
RP51D08	—	0.015	0.943 (0.133)	-0.303 (0.132)	0.153 (0.018)			
RP51D09	—	0.014	0.844 (0.078)	0.092 (0.164)	0.150 (0.014)			
RP51D10	—	0.019	0.518 (0.050)	0.495 (0.110)		-2.261 (0.332)	2.261 (0.332)	
RP51D11	—	0.024	0.606 (0.052)	0.148 (0.122)		0.542 (0.080)	-0.542 (0.080)	
RP51D12	—	0.019	0.708 (0.059)	-0.082 (0.102)		0.603 (0.089)	0.270 (0.072)	-0.873 (0.096)
RP51D13	—	0.019	0.743 (0.074)	-0.308 (0.146)				
RP51D14	—	0.019	1.024 (0.082)	0.057 (0.130)	0.173 (0.023)			
RP51D15	—	0.035	0.769 (0.061)	0.988 (0.166)		0.170 (0.168)	-0.170 (0.168)	
RP51D16	—	0.025	0.859 (0.085)	0.423 (0.158)				
RP51N01	—	0.018	0.949 (0.076)	-0.956 (0.121)				
RP51N02	—	0.026	0.669 (0.054)	0.088 (0.098)		0.164 (0.057)	-0.164 (0.057)	
RP51N03	—	0.028	0.962 (0.126)	0.796 (0.199)	0.112 (0.009)			

Paper item parameters estimated from the PIRLS 2021 concurrent calibration. Used to estimate student proficiency for paperPIRLS countries and bridge data from digitalPIRLS countries.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP51N04	—	0.024	0.716 (0.074)	-0.366 (0.119)				
RP51N05	—	0.033	0.765 (0.077)	0.619 (0.116)				
RP51N06	—	0.016	0.770 (0.046)	0.085 (0.063)		0.034 (0.061)	-0.034 (0.061)	
RP51N07	—	0.020	0.609 (0.062)	-0.504 (0.131)	0.206 (0.023)			
RP51N08	—	0.013	0.907 (0.109)	0.018 (0.144)	0.179 (0.027)			
RP51N09	—	0.022	0.827 (0.134)	0.684 (0.204)		0.054 (0.142)	-0.054 (0.142)	
RP51N10	—	0.014	1.179 (0.082)	0.607 (0.096)				
RP51N11	—	0.015	1.001 (0.088)	0.348 (0.124)				
RP51N12	—	0.022	0.734 (0.054)	0.588 (0.090)				
RP51N13	—	0.017	0.957 (0.046)	-0.094 (0.061)		0.247 (0.031)	-0.247 (0.031)	
RP51N14	—	0.019	0.911 (0.101)	0.953 (0.134)				
RP51N15	—	0.026	0.768 (0.087)	1.567 (0.174)				
RP51R01	—	0.018	0.759 (0.070)	-1.256 (0.111)	0.170 (0.022)			
RP51R02	—	0.017	0.947 (0.127)	-1.649 (0.342)				
RP51R03	—	0.012	0.763 (0.053)	-3.372 (0.259)				
RP51R04	—	0.021	0.693 (0.055)	-0.395 (0.145)				
RP51R05	—	0.012	0.619 (0.056)	-0.916 (0.128)		-0.303 (0.162)	0.303 (0.162)	
RP51R06	—	0.010	1.038 (0.064)	-1.291 (0.138)	0.234 (0.043)			
RP51R07	—	0.014	0.775 (0.053)	-1.443 (0.159)				
RP51R08	—	0.017	0.783 (0.120)	-0.869 (0.306)	0.308 (0.057)			
RP51R09	—	0.010	1.062 (0.064)	-0.387 (0.110)	0.194 (0.035)			
RP51R10	—	0.018	0.919 (0.075)	-0.168 (0.133)	0.127 (0.016)			
RP51R11	—	0.009	0.786 (0.106)	-0.836 (0.177)	0.214 (0.021)			
RP51R13	—	0.024	0.473 (0.087)	-0.366 (0.252)				
RP51R14	—	0.017	0.690 (0.096)	-1.441 (0.129)	0.210 (0.029)			
RP51R15	—	0.019	1.132 (0.069)	-0.843 (0.107)				
RP51R16	—	0.038	0.646 (0.100)	-0.418 (0.242)				
RP51R17	—	0.018	0.718 (0.085)	-0.549 (0.122)		0.606 (0.062)	-0.606 (0.062)	
RP51T01	—	0.019	0.900 (0.091)	-2.144 (0.202)				
RP51T02	—	0.024	0.766 (0.059)	0.515 (0.117)	0.175 (0.024)			
RP51T03	—	0.015	0.743 (0.079)	-0.946 (0.148)	0.197 (0.018)			
RP51T04	—	0.025	0.733 (0.056)	-0.420 (0.193)				
RP51T05	—	0.015	0.559 (0.084)	-0.160 (0.113)		-0.122 (0.092)	0.122 (0.092)	
RP51T06	—	0.017	1.219 (0.062)	-0.718 (0.073)	0.169 (0.016)			
RP51T07	—	0.014	1.183 (0.050)	-0.558 (0.071)	0.167 (0.026)			

Paper item parameters estimated from the PIRLS 2021 concurrent calibration. Used to estimate student proficiency for paperPIRLS countries and bridge data from digitalPIRLS countries.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP51T08	—	0.012	1.067 (0.051)	-1.234 (0.099)	0.196 (0.019)			
RP51T09	—	0.026	0.777 (0.063)	-0.384 (0.134)	0.198 (0.020)			
RP51T10	—	0.020	0.991 (0.057)	-0.260 (0.077)				
RP51T11	—	0.023	0.761 (0.065)	-0.718 (0.065)		-0.699 (0.124)	0.699 (0.124)	
RP51T12	—	0.034	0.789 (0.116)	-0.461 (0.151)				
RP51T13	—	0.032	0.814 (0.059)	0.488 (0.085)		0.731 (0.113)	-0.731 (0.113)	
RP51T14	—	0.014	1.014 (0.047)	0.001 (0.057)				
RP51T15	—	0.031	0.747 (0.052)	0.218 (0.078)		0.692 (0.093)	-0.692 (0.093)	
RP51T16	—	0.021	0.798 (0.067)	0.360 (0.086)		0.411 (0.087)	-0.411 (0.087)	
RP51Z01	—	0.019	0.383 (0.132)	-0.911 (0.387)		0.478 (0.153)	-0.478 (0.153)	
RP51Z02	—	0.011	0.676 (0.100)	-1.293 (0.190)				
RP51Z03	—	0.022	1.163 (0.088)	-0.289 (0.119)	0.123 (0.022)			
RP51Z04	—	0.009	1.082 (0.078)	-0.743 (0.087)	0.224 (0.037)			
RP51Z05	—	0.022	0.814 (0.133)	0.005 (0.139)				
RP51Z06	—	0.009	0.748 (0.085)	-1.596 (0.171)		0.087 (0.100)	-0.087 (0.100)	
RP51Z07	—	0.015	1.135 (0.046)	-1.398 (0.089)				
RP51Z08	—	0.017	1.212 (0.097)	-0.716 (0.166)	0.147 (0.017)			
RP51Z09	—	0.027	0.828 (0.130)	0.120 (0.161)				
RP51Z10	—	0.009	1.059 (0.046)	-0.876 (0.071)				
RP51Z11	—	0.015	0.879 (0.084)	-1.213 (0.142)				
RP51Z12	—	0.013	0.837 (0.076)	-0.689 (0.108)		0.078 (0.096)	-0.078 (0.096)	
RP51Z13	—	0.025	1.208 (0.088)	-1.035 (0.185)	0.152 (0.014)			
RP51Z14	—	0.024	0.435 (0.038)	-0.946 (0.133)		-0.425 (0.154)	-0.054 (0.127)	0.479 (0.160)
RP51Z15	—	0.021	0.856 (0.067)	-0.324 (0.083)				
Non-Response Indicators								
RPCRINF	0.032	0.015	1.579 (0.203)	0.510 (0.083)	0.519 (0.037)			
RPCRLIT	0.028	0.017	1.418 (0.146)	0.208 (0.101)	0.527 (0.041)			
RPMCINF	0.007	0.011	1.381 (0.162)	0.393 (0.084)	0.802 (0.011)			
RPMCLIT	0.018	0.012	1.263 (0.130)	-0.058 (0.118)	0.807 (0.014)			
RPCRRSI	0.026	0.019	1.739 (0.207)	0.136 (0.099)	0.574 (0.041)			
RPCRIIE	0.035	0.014	1.495 (0.162)	0.490 (0.074)	0.445 (0.036)			
RPMCRSI	0.022	0.013	1.443 (0.163)	0.192 (0.092)	0.766 (0.014)			
RPMCIIE	0.004	0.011	1.465 (0.163)	0.153 (0.093)	0.782 (0.013)			

( ) Standard errors appear in parentheses.
* Items from passages developed for PIRLS Literacy in 2011 or 2016, but treated as new in 2021.

Appendix 11B; Item Parameters from the PIRLS 2021 Digital and ePIRLS Item Calibrations

Paper item parameters estimated from PIRLS 2021 digital calibration. Used to link the paper and digital assessments.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
Items Released in 2016								
R11F01M	0.020	—	1.267 (0.080)	-0.811 (0.066)	0.109 (0.014)			
R11F02M	0.012	—	0.720 (0.051)	-1.029 (0.102)	0.219 (0.029)			
R11F03M	0.013	—	1.053 (0.067)	-0.811 (0.088)	0.157 (0.020)			
R11F04M	0.009	—	1.363 (0.053)	-0.963 (0.078)	0.174 (0.019)			
R11F05M	0.017	—	0.972 (0.064)	-0.478 (0.142)	0.191 (0.028)			
R11F06C	0.016	—	0.765 (0.056)	-0.298 (0.090)				
R11F07C	0.024	—	0.465 (0.035)	0.137 (0.082)		-1.094 (0.146)	1.094 (0.146)	
R11F08C	0.029	—	1.043 (0.041)	-0.547 (0.077)				
R11F09C	0.012	—	1.002 (0.062)	-0.817 (0.087)		0.099 (0.038)	-0.099 (0.038)	
R11F10C	0.010	—	0.797 (0.068)	-1.713 (0.176)				
R11F11M	0.013	—	0.650 (0.066)	-0.085 (0.114)	0.185 (0.024)			
R11F12C	0.019	—	0.643 (0.032)	0.346 (0.068)		-0.397 (0.094)	0.397 (0.094)	
R11F13M	0.011	—	1.028 (0.068)	-0.520 (0.101)	0.199 (0.036)			
R11L01M	0.023	—	0.616 (0.053)	-2.155 (0.171)	0.181 (0.020)			
R11L02M	0.019	—	0.678 (0.055)	0.293 (0.106)	0.189 (0.028)			
R11L03C	0.015	—	0.647 (0.042)	-0.674 (0.093)				
R11L04C	0.042	—	0.724 (0.039)	0.124 (0.065)		1.937 (0.110)	-1.151 (0.105)	-0.787 (0.064)
R11L05M	0.018	—	1.093 (0.073)	0.249 (0.080)	0.189 (0.020)			
R11L06C	0.017	—	0.632 (0.041)	0.080 (0.089)				
R11L07M	0.024	—	0.726 (0.070)	0.314 (0.094)	0.158 (0.016)			
R11L08C	0.022	—	0.819 (0.043)	0.267 (0.072)		0.847 (0.056)	-0.847 (0.056)	
R11L09M	0.009	—	0.999 (0.066)	-1.121 (0.129)	0.209 (0.028)			
R11L10C	0.024	—	0.693 (0.038)	0.409 (0.095)		0.393 (0.074)	-0.393 (0.074)	
R11L11M	0.014	—	0.919 (0.051)	-0.508 (0.099)	0.187 (0.024)			
R11L12C	0.018	—	0.714 (0.053)	0.181 (0.091)		1.050 (0.091)	-1.050 (0.091)	
R41H01M	0.009	—	0.844 (0.082)	-1.242 (0.359)	0.457 (0.169)			
R41H02M	0.011	—	1.047 (0.047)	-0.931 (0.086)	0.192 (0.036)			

Paper item parameters estimated from PIRLS 2021 digital calibration. Used to link the paper and digital assessments.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
R41H03C	0.012	—	1.167 (0.069)	-0.016 (0.056)				
R41H04C	0.024	—	0.664 (0.058)	1.379 (0.158)				
R41H05M	0.009	—	0.911 (0.093)	-0.056 (0.095)	0.277 (0.029)			
R41H06C	0.011	—	0.699 (0.040)	-0.591 (0.089)		0.093 (0.050)	-0.093 (0.050)	
R41H07M	0.027	—	0.832 (0.058)	0.585 (0.093)	0.127 (0.012)			
R41H08C	0.026	—	0.690 (0.042)	1.229 (0.102)				
R41H09M	0.017	—	0.691 (0.063)	-0.013 (0.099)	0.169 (0.030)			
R41H10M	0.013	—	1.247 (0.048)	-0.432 (0.067)	0.222 (0.027)			
R41H11M	0.007	—	1.274 (0.052)	-0.865 (0.085)	0.259 (0.048)			
R41H12M	0.010	—	1.250 (0.051)	-0.293 (0.076)	0.324 (0.022)			
R41H13C	0.017	—	0.519 (0.031)	0.520 (0.073)		0.154 (0.124)	0.163 (0.114)	-0.317 (0.074)
R41H14C	0.031	—	0.898 (0.053)	-0.051 (0.072)				
R41H15C	0.015	—	1.225 (0.045)	-0.654 (0.078)				
R41H16C	0.022	—	0.937 (0.053)	0.382 (0.083)				
R41T01M	0.006	—	1.080 (0.047)	-1.270 (0.092)	0.229 (0.032)			
R41T02C	0.016	—	0.699 (0.047)	-0.827 (0.092)		0.380 (0.039)	-0.380 (0.039)	
R41T03C	0.011	—	0.972 (0.035)	-0.187 (0.065)		0.079 (0.022)	-0.079 (0.022)	
R41T04C	0.008	—	1.112 (0.046)	-0.327 (0.074)				
R41T05M	0.025	—	0.774 (0.077)	0.164 (0.143)	0.254 (0.052)			
R41T06C	0.010	—	1.346 (0.060)	-0.955 (0.084)				
R41T07C	0.020	—	0.759 (0.032)	0.415 (0.065)		-0.217 (0.044)	0.217 (0.044)	
R41T08C	0.014	—	1.130 (0.041)	-0.313 (0.077)				
R41T09M	0.021	—	1.301 (0.069)	0.211 (0.065)	0.149 (0.019)			
R41T10C	0.006	—	1.247 (0.048)	-0.569 (0.075)				
R41T11C	0.019	—	0.734 (0.035)	0.288 (0.061)		-0.392 (0.067)	0.309 (0.048)	0.083 (0.057)
R41T12M	0.010	—	0.937 (0.059)	-0.537 (0.152)	0.271 (0.091)			
R41T13M	0.012	—	1.118 (0.055)	0.194 (0.067)	0.113 (0.009)			
R41T14C	0.019	—	0.455 (0.035)	-0.194 (0.119)				
R41T15M	0.013	—	0.879 (0.050)	0.039 (0.085)	0.207 (0.028)			
R41T16M	0.018	—	1.148 (0.069)	-0.399 (0.078)	0.210 (0.028)			
Items Common in 2016 and 2021								
RP21K01	0.015	0.020	0.468 (0.036)	-1.035 (0.107)		0.239 (0.050)	-0.239 (0.050)	
RP21K02	0.027	0.023	0.772 (0.046)	-0.924 (0.101)				
RP21K03	0.021	0.010	0.892 (0.061)	-0.273 (0.089)	0.159 (0.026)			
RP21K05	0.018	0.020	1.039 (0.048)	-0.051 (0.090)				

Paper item parameters estimated from PIRLS 2021 digital calibration. Used to link the paper and digital assessments.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP21K06	0.011	0.019	1.406 (0.087)	-0.252 (0.084)	0.267 (0.021)			
RP21K07	0.026	0.026	0.781 (0.044)	-0.125 (0.084)		0.227 (0.032)	-0.227 (0.032)	
RP21K08	0.023	0.018	0.973 (0.071)	0.085 (0.090)	0.154 (0.030)			
RP21K09	0.013	0.018	1.249 (0.067)	-0.308 (0.092)	0.202 (0.021)			
RP21K10	0.013	0.021	0.690 (0.037)	0.421 (0.058)		-0.286 (0.088)	0.286 (0.088)	
RP21K11	0.015	0.013	1.025 (0.092)	-0.197 (0.084)	0.169 (0.027)			
RP21K12	0.021	0.023	0.591 (0.032)	-0.468 (0.078)		0.854 (0.068)	-0.269 (0.052)	-0.585 (0.047)
RP21Y01	0.016	0.012	1.098 (0.060)	-0.062 (0.079)	0.220 (0.024)			
RP21Y02	0.013	0.007	1.577 (0.065)	-0.485 (0.077)	0.237 (0.024)			
RP21Y03	0.044	0.031	0.635 (0.053)	0.252 (0.094)				
RP21Y04	0.016	0.015	1.197 (0.075)	-0.198 (0.072)	0.168 (0.022)			
RP21Y05	0.009	0.014	1.610 (0.076)	-0.145 (0.075)	0.210 (0.017)			
RP21Y06	0.017	0.011	1.540 (0.063)	-0.174 (0.061)	0.188 (0.018)			
RP21Y07	0.028	0.018	0.854 (0.064)	-1.117 (0.090)	0.131 (0.020)			
RP21Y08	0.024	0.007	1.487 (0.050)	-0.396 (0.078)	0.228 (0.029)			
RP21Y09	0.017	0.013	1.013 (0.045)	-0.804 (0.083)		0.220 (0.024)	-0.220 (0.024)	
RP21Y10	0.033	0.026	0.669 (0.033)	0.306 (0.072)				
RP21Y11	0.012	0.017	1.312 (0.073)	-0.235 (0.071)	0.276 (0.020)			
RP21Y12	0.016	0.014	0.701 (0.030)	-0.164 (0.071)		-1.134 (0.123)	1.134 (0.123)	
RP21Y13	0.030	0.016	0.762 (0.033)	0.054 (0.075)		0.973 (0.074)	-0.474 (0.081)	-0.499 (0.049)
RP21Y14	0.023	0.020	0.551 (0.028)	0.057 (0.070)		-0.556 (0.094)	0.556 (0.094)	
RP31M01	0.011	0.014	1.420 (0.082)	-1.160 (0.105)	0.168 (0.027)			
RP31M02	0.007	0.009	1.224 (0.054)	-1.292 (0.108)				
RP31M03	0.011	0.009	1.204 (0.059)	-0.363 (0.070)	0.203 (0.031)			
RP31M04	0.019	0.019	0.577 (0.053)	0.021 (0.107)				
RP31M05	0.019	0.025	1.293 (0.109)	-0.156 (0.073)	0.388 (0.026)			
RP31M06	0.018	0.023	1.097 (0.074)	-0.024 (0.097)	0.256 (0.020)			
RP31M07	0.024	0.010	1.457 (0.082)	-0.591 (0.070)	0.185 (0.020)			
RP31M08	0.015	0.013	1.350 (0.050)	-0.806 (0.085)	0.187 (0.032)			
RP31M09	0.037	0.032	0.682 (0.038)	-0.486 (0.072)		1.423 (0.095)	-1.423 (0.095)	
RP31M10	0.044	0.041	0.570 (0.055)	0.027 (0.128)				
RP31M11	0.028	0.012	0.863 (0.065)	-1.021 (0.133)	0.205 (0.040)			
RP31M12	0.019	0.020	1.153 (0.055)	-0.242 (0.076)	0.127 (0.017)			
RP31M13	0.010	0.011	1.882 (0.062)	-1.004 (0.088)	0.190 (0.021)			
RP31M14	0.026	0.017	2.003 (0.073)	-0.554 (0.074)	0.160 (0.013)			

Paper item parameters estimated from PIRLS 2021 digital calibration. Used to link the paper and digital assessments.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP31M15	0.018	0.019	1.248 (0.085)	-0.357 (0.084)	0.196 (0.016)			
RP31M16	0.026	0.029	1.207 (0.048)	-0.289 (0.069)				
RP31M17	0.021	0.020	0.581 (0.032)	-0.381 (0.073)		0.361 (0.058)	0.100 (0.036)	-0.460 (0.044)
RP31W01	0.037	0.038	0.796 (0.038)	-0.768 (0.076)		0.240 (0.038)	-0.240 (0.038)	
RP31W02	0.019	0.027	0.781 (0.033)	0.005 (0.065)		-0.090 (0.037)	0.090 (0.037)	
RP31W03	0.020	0.010	1.326 (0.105)	-0.267 (0.086)	0.154 (0.026)			
RP31W04	0.022	0.013	0.864 (0.052)	-0.937 (0.090)				
RP31W05	0.024	0.010	1.185 (0.071)	0.166 (0.079)	0.247 (0.021)			
RP31W06	0.021	0.011	0.771 (0.043)	-1.204 (0.137)	0.174 (0.029)			
RP31W07	0.020	0.019	0.932 (0.039)	0.258 (0.082)		0.017 (0.031)	0.115 (0.028)	-0.132 (0.029)
RP31W08	0.006	0.011	1.279 (0.062)	-0.432 (0.090)	0.243 (0.026)			
RP31W09	0.026	0.026	0.859 (0.106)	0.161 (0.089)	0.136 (0.024)			
RP31W10	0.012	0.016	1.285 (0.071)	0.064 (0.073)	0.166 (0.015)			
RP31W11	0.020	0.019	1.384 (0.053)	0.224 (0.075)				
RP31W12	0.025	0.011	1.462 (0.064)	0.395 (0.076)	0.200 (0.012)			
RP31W13	0.030	0.021	0.794 (0.036)	0.456 (0.075)				
RP41B01	0.008	0.014	0.752 (0.056)	-2.512 (0.187)				
RP41B02	0.016	0.013	1.076 (0.059)	-2.172 (0.121)	0.158 (0.024)			
RP41B03	0.018	0.017	0.747 (0.063)	-2.558 (0.199)	0.169 (0.028)			
RP41B04	0.014	0.011	0.807 (0.041)	-1.258 (0.098)				
RP41B05	0.014	0.007	0.995 (0.046)	-1.473 (0.092)	0.164 (0.028)			
RP41B06	0.018	0.022	0.551 (0.050)	-2.046 (0.242)	0.176 (0.031)			
RP41B07	0.019	0.013	0.801 (0.049)	-1.266 (0.123)				
RP41B08	0.017	0.007	0.848 (0.053)	-1.625 (0.137)				
RP41B09	0.013	0.005	1.017 (0.045)	-1.167 (0.099)				
RP41B10	0.010	0.018	0.657 (0.046)	-0.641 (0.146)	0.188 (0.030)			
RP41B11	0.016	0.007	1.029 (0.064)	-1.332 (0.149)	0.215 (0.043)			
RP41B12	0.012	0.010	0.848 (0.063)	-2.125 (0.148)	0.188 (0.039)			
RP41B13	0.025	0.036	0.506 (0.034)	-1.346 (0.131)		0.837 (0.071)	-0.837 (0.071)	
RP41B14	0.024	0.019	1.043 (0.060)	-0.900 (0.090)	0.139 (0.018)			
RP41B15	0.027	0.028	0.357 (0.028)	0.083 (0.108)				
RP41B16	0.023	0.022	0.431 (0.027)	-0.413 (0.088)		0.146 (0.051)	-0.146 (0.051)	
RP41B17	0.022	0.021	0.722 (0.037)	-0.653 (0.082)		0.893 (0.064)	-0.893 (0.064)	
RP41E01	0.017	0.013	0.726 (0.067)	-3.385 (0.184)				
RP41E02	0.023	0.021	0.801 (0.055)	-1.926 (0.187)				

Paper item parameters estimated from PIRLS 2021 digital calibration. Used to link the paper and digital assessments.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP41E03	0.015	0.024	1.297 (0.064)	-0.902 (0.091)	0.257 (0.039)			
RP41E04	0.006	0.010	1.070 (0.058)	-0.679 (0.091)	0.211 (0.026)			
RP41E05	0.010	0.009	1.261 (0.053)	-1.653 (0.093)	0.179 (0.039)			
RP41E06	0.012	0.023	1.063 (0.073)	-2.422 (0.184)	0.274 (0.072)			
RP41E07	0.024	0.021	0.516 (0.024)	-1.060 (0.096)		0.104 (0.063)	-0.104 (0.063)	
RP41E08	0.017	0.009	0.972 (0.064)	-1.128 (0.140)	0.181 (0.034)			
RP41E09	0.024	0.020	0.580 (0.116)	0.268 (0.276)	0.360 (0.067)			
RP41E10	0.025	0.018	0.822 (0.064)	-3.252 (0.186)				
RP41E11	0.019	0.006	1.365 (0.063)	-0.970 (0.098)	0.237 (0.026)			
RP41E12	0.021	0.012	1.113 (0.060)	-2.616 (0.163)				
RP41E13	0.023	0.015	0.521 (0.037)	-0.514 (0.101)				
RP41E14	0.030	0.029	0.461 (0.029)	-0.227 (0.106)		1.024 (0.078)	-1.024 (0.078)	
RP41E15	0.014	0.012	0.770 (0.035)	-2.592 (0.126)				
RP41E16	0.024	0.012	0.786 (0.038)	-1.127 (0.102)				
RP41E17	0.012	0.011	1.125 (0.067)	-0.720 (0.086)	0.148 (0.019)			
RP41I01	0.016	0.008	0.945 (0.050)	-1.570 (0.124)				
RP41I02	0.037	0.014	0.975 (0.075)	0.274 (0.090)	0.153 (0.018)			
RP41I03	0.024	0.026	0.543 (0.025)	-0.042 (0.071)		0.410 (0.031)	-0.410 (0.031)	
RP41I04	0.025	0.031	0.709 (0.027)	0.119 (0.068)		0.472 (0.043)	-0.472 (0.043)	
RP41I05	0.025	0.016	1.087 (0.064)	-0.175 (0.090)	0.199 (0.036)			
RP41I06	0.028	0.015	1.220 (0.073)	0.230 (0.100)	0.230 (0.027)			
RP41I07	0.017	0.011	0.971 (0.035)	0.197 (0.055)		0.039 (0.034)	-0.039 (0.034)	
RP41I08	0.030	0.014	1.585 (0.072)	-0.786 (0.080)	0.230 (0.036)			
RP41I09	0.027	0.008	1.151 (0.059)	-0.318 (0.072)				
RP41I10	0.010	0.022	0.856 (0.078)	-0.217 (0.106)	0.173 (0.031)			
RP41I11	0.022	0.019	0.874 (0.041)	0.165 (0.066)		0.439 (0.038)	-0.439 (0.038)	
RP41I12	0.018	0.015	1.103 (0.051)	-0.030 (0.069)	0.132 (0.018)			
RP41I13	0.018	0.018	0.873 (0.041)	-0.048 (0.081)				
RP41I14	0.015	0.012	0.798 (0.051)	0.612 (0.060)				
RP41I15	0.024	0.007	0.870 (0.045)	0.119 (0.073)				
RP41O01	0.010	0.011	0.980 (0.046)	-1.183 (0.101)	0.220 (0.032)			
RP41O02	0.010	0.009	0.937 (0.058)	-1.010 (0.077)				
RP41O03	0.034	0.028	0.923 (0.056)	0.641 (0.087)				
RP41O04	0.027	0.033	0.547 (0.049)	0.919 (0.103)		-0.049 (0.073)	0.049 (0.073)	
RP41O05	0.026	0.030	0.608 (0.032)	-0.034 (0.093)		-0.211 (0.073)	0.211 (0.073)	

Paper item parameters estimated from PIRLS 2021 digital calibration. Used to link the paper and digital assessments.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP41O06	0.005	0.012	1.422 (0.061)	-0.492 (0.085)	0.266 (0.022)			
RP41O07	0.030	0.031	0.613 (0.035)	-1.291 (0.098)		0.215 (0.039)	-0.215 (0.039)	
RP41O08	0.010	0.024	0.780 (0.049)	-0.328 (0.098)				
RP41O09	0.020	0.035	1.139 (0.049)	-0.371 (0.067)				
RP41O10	0.022	0.021	0.797 (0.035)	0.094 (0.070)		-0.023 (0.039)	0.023 (0.039)	
RP41O11	0.020	0.014	1.374 (0.058)	-0.103 (0.072)	0.260 (0.026)			
RP41O12	0.023	0.013	1.337 (0.065)	-0.525 (0.071)	0.140 (0.018)			
RP41O13	0.018	0.019	0.542 (0.041)	0.120 (0.063)		-1.173 (0.222)	0.908 (0.159)	0.265 (0.096)
Items Introduced in 2021								
* RP31P01	—	0.009	0.855 (0.116)	-0.948 (0.209)	0.238 (0.057)			
* RP31P02	—	0.006	1.018 (0.050)	-1.715 (0.198)	0.201 (0.016)			
* RP31P03	—	0.019	0.795 (0.074)	-1.126 (0.130)		0.504 (0.078)	-0.504 (0.078)	
* RP31P04	—	0.010	0.936 (0.044)	-2.300 (0.189)				
* RP31P05	—	0.018	1.213 (0.070)	-1.413 (0.178)	0.197 (0.027)			
* RP31P06	—	0.015	0.646 (0.095)	-2.714 (0.308)				
* RP31P07	—	0.027	0.743 (0.070)	-0.901 (0.116)				
* RP31P08	—	0.020	1.055 (0.099)	-1.574 (0.228)	0.201 (0.029)			
* RP31P09	—	0.020	1.213 (0.062)	-1.621 (0.135)	0.221 (0.026)			
* RP31P10	—	0.011	1.087 (0.042)	-1.631 (0.128)	0.197 (0.015)			
* RP31P11	—	0.016	0.755 (0.065)	-1.264 (0.145)		-0.261 (0.141)	0.261 (0.141)	
* RP31P12	—	0.031	0.577 (0.118)	-0.664 (0.244)				
* RP31P13	—	0.014	1.111 (0.053)	-1.307 (0.125)				
* RP31P14	—	0.012	1.030 (0.047)	-0.872 (0.092)				
* RP31U01	—	0.006	0.729 (0.060)	-2.691 (0.205)				
* RP31U02	—	0.011	0.886 (0.060)	-2.516 (0.150)	0.257 (0.017)			
* RP31U03	—	0.009	0.964 (0.060)	-1.708 (0.138)	0.202 (0.020)			
* RP31U04	—	0.018	0.625 (0.102)	-1.700 (0.304)				
* RP31U05	—	0.017	0.930 (0.066)	-1.661 (0.162)	0.206 (0.035)			
* RP31U06	—	0.014	1.033 (0.066)	-1.645 (0.137)	0.186 (0.021)			
* RP31U07	—	0.014	1.079 (0.067)	-0.626 (0.103)	0.179 (0.020)			
* RP31U08	—	0.010	1.003 (0.095)	-0.791 (0.146)	0.191 (0.022)			
* RP31U09	—	0.008	1.059 (0.045)	-1.690 (0.100)	0.221 (0.025)			
* RP31U10	—	0.018	0.734 (0.092)	-0.753 (0.167)				
* RP31U11	—	0.016	0.725 (0.052)	-1.723 (0.098)		0.204 (0.119)	-0.204 (0.119)	
* RP31U12	—	0.018	1.058 (0.064)	-1.628 (0.145)				

Paper item parameters estimated from PIRLS 2021 digital calibration. Used to link the paper and digital assessments.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
* RP31U13	—	0.021	1.147 (0.047)	-1.796 (0.096)	0.173 (0.013)			
* RP31U14	—	0.049	0.538 (0.100)	0.053 (0.171)				
* RP41H01	—	0.008	0.864 (0.043)	-2.665 (0.165)				
* RP41H02	—	0.017	1.065 (0.051)	-1.714 (0.108)	0.201 (0.018)			
* RP41H03	—	0.026	0.701 (0.089)	-0.353 (0.177)	0.198 (0.028)			
* RP41H04	—	0.020	0.744 (0.083)	-2.049 (0.177)	0.219 (0.019)			
* RP41H05	—	0.021	0.974 (0.093)	-2.413 (0.319)	0.271 (0.025)			
* RP41H06	—	0.014	0.869 (0.110)	-2.330 (0.248)				
* RP41H07	—	0.013	1.138 (0.053)	-1.193 (0.099)	0.204 (0.024)			
* RP41H08	—	0.008	1.056 (0.063)	-1.044 (0.129)	0.169 (0.015)			
* RP41H09	—	0.019	1.035 (0.055)	-1.682 (0.113)	0.206 (0.022)			
* RP41H10	—	0.009	0.976 (0.062)	-1.099 (0.103)	0.212 (0.031)			
* RP41H11	—	0.011	0.693 (0.072)	-1.857 (0.196)				
* RP41H12	—	0.022	0.969 (0.064)	-1.286 (0.123)				
* RP41H13	—	0.008	1.123 (0.051)	-1.396 (0.110)	0.180 (0.026)			
* RP41H14	—	0.026	0.857 (0.136)	-1.982 (0.142)				
* RP41H15	—	0.014	0.939 (0.057)	-0.728 (0.143)	0.177 (0.024)			
* RP41H16	—	0.011	1.097 (0.046)	-1.080 (0.110)	0.194 (0.020)			
* RP41M01	—	0.017	0.631 (0.148)	-2.724 (0.548)	0.299 (0.046)			
* RP41M02	—	0.014	1.111 (0.052)	-0.445 (0.132)	0.175 (0.028)			
* RP41M03	—	0.015	0.666 (0.056)	-2.347 (0.184)				
* RP41M04	—	0.015	0.861 (0.108)	-1.289 (0.122)				
* RP41M05	—	0.012	0.807 (0.066)	-2.456 (0.198)	0.232 (0.020)			
* RP41M06	—	0.009	0.923 (0.157)	-2.119 (0.628)				
* RP41M07	—	0.007	0.962 (0.062)	-1.321 (0.156)	0.237 (0.030)			
* RP41M08	—	0.022	1.034 (0.067)	-2.391 (0.194)				
* RP41M09	—	0.006	0.961 (0.058)	-1.518 (0.111)	0.216 (0.036)			
* RP41M10	—	0.008	0.715 (0.104)	-1.995 (0.253)		0.184 (0.085)	-0.184 (0.085)	
* RP41M11	—	0.008	1.094 (0.055)	-0.717 (0.116)	0.186 (0.039)			
* RP41M12	—	0.015	0.799 (0.174)	-1.992 (0.316)				
* RP41M13	—	0.022	1.012 (0.086)	-2.411 (0.243)	0.210 (0.036)			
* RP41M14	—	0.020	0.653 (0.039)	-0.849 (0.126)		0.665 (0.144)	0.064 (0.082)	-0.729 (0.093)
* RP41M15	—	0.009	1.043 (0.079)	-1.021 (0.137)	0.215 (0.034)			
* RP41M16	—	0.009	0.888 (0.090)	-1.750 (0.200)				
* RP41M17	—	0.013	1.141 (0.094)	-0.695 (0.120)				

Paper item parameters estimated from PIRLS 2021 digital calibration. Used to link the paper and digital assessments.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
* RP41M18	—	0.021	1.141 (0.111)	-0.987 (0.125)	0.152 (0.019)			
RP51C01	—	0.019	0.468 (0.107)	0.259 (0.123)		0.095 (0.086)	-0.095 (0.086)	
RP51C02	—	0.010	0.992 (0.057)	-0.979 (0.096)	0.172 (0.022)			
RP51C03	—	0.020	0.987 (0.057)	0.317 (0.115)				
RP51C04	—	0.008	0.982 (0.062)	-1.350 (0.142)	0.206 (0.030)			
RP51C05	—	0.018	0.993 (0.047)	-0.862 (0.081)		0.368 (0.046)	-0.368 (0.046)	
RP51C06	—	0.024	0.605 (0.066)	0.189 (0.144)		-0.064 (0.158)	0.064 (0.158)	
RP51C07	—	0.011	0.709 (0.090)	-0.677 (0.146)				
RP51C08	—	0.013	0.906 (0.062)	-0.420 (0.121)	0.186 (0.029)			
RP51C09	—	0.023	1.227 (0.062)	-1.218 (0.121)	0.166 (0.016)			
RP51C10	—	0.024	0.848 (0.052)	-0.436 (0.085)		0.651 (0.076)	-0.651 (0.076)	
RP51C11	—	0.017	0.966 (0.063)	-0.478 (0.117)	0.165 (0.018)			
RP51C12	—	0.010	1.197 (0.056)	-0.845 (0.092)	0.196 (0.026)			
RP51C13	—	0.028	0.878 (0.054)	-0.225 (0.096)		0.356 (0.056)	-0.356 (0.056)	
RP51C14	—	0.034	0.947 (0.067)	0.626 (0.087)		0.173 (0.071)	-0.173 (0.071)	
RP51C15	—	0.030	0.780 (0.103)	0.960 (0.144)				
RP51D01	—	0.006	0.940 (0.076)	-1.247 (0.128)				
RP51D02	—	0.023	0.795 (0.078)	-0.071 (0.129)	0.210 (0.037)			
RP51D03	—	0.014	0.902 (0.058)	-0.392 (0.099)				
RP51D04	—	0.011	1.073 (0.100)	-0.543 (0.135)	0.156 (0.018)			
RP51D05	—	0.017	0.898 (0.064)	-0.540 (0.108)	0.186 (0.021)			
RP51D06	—	0.033	0.710 (0.051)	-0.315 (0.091)		0.628 (0.064)	-0.628 (0.064)	
RP51D07	—	0.014	1.031 (0.084)	0.147 (0.096)				
RP51D08	—	0.015	0.967 (0.132)	-0.500 (0.135)	0.160 (0.019)			
RP51D09	—	0.016	0.878 (0.081)	-0.106 (0.174)	0.158 (0.016)			
RP51D10	—	0.020	0.526 (0.051)	0.269 (0.102)		-2.224 (0.325)	2.224 (0.325)	
RP51D11	—	0.025	0.615 (0.052)	-0.073 (0.120)		0.535 (0.079)	-0.535 (0.079)	
RP51D12	—	0.018	0.718 (0.059)	-0.299 (0.109)		0.596 (0.089)	0.265 (0.070)	-0.861 (0.096)
RP51D13	—	0.020	0.761 (0.075)	-0.514 (0.137)				
RP51D14	—	0.018	1.062 (0.083)	-0.142 (0.134)	0.181 (0.025)			
RP51D15	—	0.044	0.781 (0.062)	0.755 (0.165)		0.168 (0.166)	-0.168 (0.166)	
RP51D16	—	0.030	0.885 (0.087)	0.201 (0.157)				
RP51N01	—	0.014	0.955 (0.077)	-1.165 (0.127)				
RP51N02	—	0.029	0.677 (0.054)	-0.134 (0.097)		0.162 (0.057)	-0.162 (0.057)	
RP51N03	—	0.031	1.013 (0.128)	0.570 (0.188)	0.117 (0.009)			

Paper item parameters estimated from PIRLS 2021 digital calibration. Used to link the paper and digital assessments.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP51N04	—	0.023	0.730 (0.074)	-0.576 (0.125)				
RP51N05	—	0.034	0.790 (0.080)	0.389 (0.122)				
RP51N06	—	0.016	0.779 (0.047)	-0.138 (0.065)		0.033 (0.060)	-0.033 (0.060)	
RP51N07	—	0.021	0.635 (0.066)	-0.663 (0.148)	0.223 (0.030)			
RP51N08	—	0.012	0.943 (0.110)	-0.176 (0.144)	0.189 (0.030)			
RP51N09	—	0.027	0.839 (0.134)	0.454 (0.204)		0.053 (0.140)	-0.053 (0.140)	
RP51N10	—	0.017	1.215 (0.083)	0.379 (0.100)				
RP51N11	—	0.020	1.030 (0.089)	0.127 (0.117)				
RP51N12	—	0.029	0.759 (0.056)	0.358 (0.082)				
RP51N13	—	0.019	0.967 (0.046)	-0.314 (0.076)		0.244 (0.031)	-0.244 (0.031)	
RP51N14	—	0.027	0.951 (0.105)	0.712 (0.129)				
RP51N15	—	0.038	0.818 (0.093)	1.286 (0.165)				
RP51R01	—	0.016	0.762 (0.068)	-1.448 (0.112)	0.176 (0.023)			
RP51R02	—	0.016	0.935 (0.128)	-1.863 (0.363)				
RP51R03	—	0.011	0.735 (0.049)	-3.646 (0.270)				
RP51R04	—	0.023	0.703 (0.056)	-0.599 (0.150)				
RP51R05	—	0.011	0.621 (0.058)	-1.124 (0.140)		-0.303 (0.162)	0.303 (0.162)	
RP51R06	—	0.009	1.032 (0.065)	-1.488 (0.147)	0.241 (0.047)			
RP51R07	—	0.010	0.774 (0.052)	-1.649 (0.170)				
RP51R08	—	0.015	0.812 (0.116)	-1.008 (0.305)	0.331 (0.061)			
RP51R09	—	0.013	1.082 (0.062)	-0.577 (0.109)	0.201 (0.038)			
RP51R10	—	0.018	0.940 (0.073)	-0.365 (0.141)	0.132 (0.017)			
RP51R11	—	0.010	0.801 (0.107)	-1.011 (0.186)	0.226 (0.024)			
RP51R13	—	0.028	0.482 (0.089)	-0.571 (0.244)				
RP51R14	—	0.016	0.696 (0.094)	-1.619 (0.142)	0.220 (0.033)			
RP51R15	—	0.015	1.130 (0.069)	-1.048 (0.113)				
RP51R16	—	0.042	0.656 (0.100)	-0.620 (0.234)				
RP51R17	—	0.019	0.724 (0.085)	-0.758 (0.125)		0.602 (0.061)	-0.602 (0.061)	
RP51T01	—	0.018	0.884 (0.088)	-2.366 (0.224)				
RP51T02	—	0.028	0.822 (0.060)	0.320 (0.111)	0.187 (0.028)			
RP51T03	—	0.015	0.760 (0.080)	-1.117 (0.146)	0.208 (0.021)			
RP51T04	—	0.024	0.748 (0.057)	-0.621 (0.184)				
RP51T05	—	0.014	0.567 (0.086)	-0.371 (0.106)		-0.120 (0.090)	0.120 (0.090)	
RP51T06	—	0.018	1.227 (0.060)	-0.913 (0.082)	0.174 (0.017)			
RP51T07	—	0.013	1.199 (0.051)	-0.752 (0.085)	0.173 (0.028)			

Paper item parameters estimated from PIRLS 2021 digital calibration. Used to link the paper and digital assessments.

Item	RMSD 2016	RMSD 2021	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RP51T08	—	0.012	1.065 (0.049)	-1.428 (0.105)	0.202 (0.020)			
RP51T09	—	0.028	0.805 (0.065)	-0.559 (0.138)	0.209 (0.023)			
RP51T10	—	0.020	1.009 (0.059)	-0.464 (0.077)				
RP51T11	—	0.021	0.770 (0.065)	-0.923 (0.088)		-0.691 (0.122)	0.691 (0.122)	
RP51T12	—	0.033	0.804 (0.115)	-0.662 (0.149)				
RP51T13	—	0.036	0.826 (0.060)	0.268 (0.087)		0.720 (0.111)	-0.720 (0.111)	
RP51T14	—	0.016	1.037 (0.049)	-0.207 (0.066)				
RP51T15	—	0.037	0.758 (0.053)	0.002 (0.080)		0.682 (0.091)	-0.682 (0.091)	
RP51T16	—	0.023	0.809 (0.067)	0.142 (0.093)		0.405 (0.086)	-0.405 (0.086)	
RP51Z01	—	0.019	0.386 (0.133)	-1.118 (0.378)		0.475 (0.152)	-0.475 (0.152)	
RP51Z02	—	0.009	0.679 (0.100)	-1.495 (0.198)				
RP51Z03	—	0.024	1.179 (0.087)	-0.488 (0.123)	0.126 (0.022)			
RP51Z04	—	0.009	1.091 (0.077)	-0.932 (0.102)	0.232 (0.040)			
RP51Z05	—	0.025	0.830 (0.138)	-0.203 (0.136)				
RP51Z06	—	0.007	0.751 (0.086)	-1.802 (0.172)		0.088 (0.100)	-0.088 (0.100)	
RP51Z07	—	0.014	1.120 (0.046)	-1.609 (0.105)				
RP51Z08	—	0.016	1.214 (0.096)	-0.915 (0.172)	0.151 (0.017)			
RP51Z09	—	0.032	0.845 (0.134)	-0.091 (0.153)				
RP51Z10	—	0.007	1.060 (0.045)	-1.080 (0.083)				
RP51Z11	—	0.012	0.880 (0.084)	-1.417 (0.147)				
RP51Z12	—	0.013	0.842 (0.075)	-0.896 (0.116)		0.077 (0.095)	-0.077 (0.095)	
RP51Z13	—	0.024	1.200 (0.089)	-1.237 (0.200)	0.156 (0.014)			
RP51Z14	—	0.022	0.439 (0.038)	-1.152 (0.141)		-0.419 (0.151)	-0.055 (0.126)	0.475 (0.157)
RP51Z15	—	0.022	0.869 (0.067)	-0.529 (0.082)				
Non-Response Indicators								
RPCRINF	0.032	0.015	1.606 (0.200)	0.284 (0.079)	0.520 (0.036)			
RPCRLIT	0.028	0.017	1.434 (0.147)	-0.015 (0.098)	0.527 (0.041)			
RPMCINF	0.007	0.011	1.400 (0.161)	0.173 (0.080)	0.803 (0.011)			
RPMCLIT	0.018	0.012	1.275 (0.128)	-0.281 (0.115)	0.807 (0.014)			
RPCRRSI	0.026	0.019	1.760 (0.218)	-0.085 (0.100)	0.575 (0.042)			
RPCRIIE	0.034	0.014	1.518 (0.169)	0.264 (0.073)	0.445 (0.036)			
RPMCRSI	0.022	0.013	1.461 (0.169)	-0.030 (0.093)	0.767 (0.014)			
RPMCIIE	0.004	0.011	1.474 (0.168)	-0.070 (0.094)	0.782 (0.013)			

( ) Standard errors appear in parentheses.
* Items from passages developed for PIRLS Literacy in 2011 or 2016, but treated as new in 2021.

digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for the digitalPIRLS countries, along with the ePIRLS items.

Item	RMSD	Slope (a$_i$)	Location (b$_i$)	Guessing (c$_i$)	Step 1 (t$_{i1}$)	Step 2 (t$_{i2}$)	Step 3 (t$_{i3}$)
			digitalPIRLS Items				
RE21K01	0.016	0.462 (0.037)	-1.533 (0.157)		0.232 (0.090)	-0.232 (0.090)	
RE21K02	0.019	0.728 (0.051)	-0.955 (0.124)				
RE21K03	0.021	0.852 (0.060)	-0.373 (0.115)	0.161 (0.018)			
RE21K05	0.012	1.257 (0.066)	0.081 (0.084)				
RE21K06	0.025	1.312 (0.068)	0.096 (0.080)	0.222 (0.021)			
RE21K07	0.020	0.972 (0.053)	0.015 (0.063)		0.046 (0.051)	-0.046 (0.051)	
RE21K08	0.015	1.015 (0.059)	0.350 (0.092)	0.120 (0.034)			
RE21K09	0.022	1.116 (0.050)	-0.170 (0.088)	0.236 (0.029)			
RE21K10	0.022	0.682 (0.057)	0.713 (0.089)		0.135 (0.126)	-0.135 (0.126)	
RE21K11	0.015	1.011 (0.071)	-0.011 (0.083)	0.153 (0.021)			
RE21K12	0.015	0.583 (0.034)	-0.275 (0.072)		0.426 (0.087)	-0.164 (0.079)	-0.263 (0.077)
RE21Y01	0.021	0.943 (0.071)	-0.072 (0.101)	0.202 (0.019)			
RE21Y02	0.009	1.337 (0.052)	-0.589 (0.070)	0.155 (0.020)			
RE21Y03	0.028	0.611 (0.049)	0.366 (0.112)				
RE21Y04	0.019	1.183 (0.075)	-0.145 (0.087)	0.153 (0.021)			
RE21Y05	0.020	1.403 (0.058)	-0.299 (0.086)	0.163 (0.018)			
RE21Y06	0.020	1.398 (0.067)	-0.311 (0.062)	0.156 (0.019)			
RE21Y07	0.032	0.804 (0.068)	-0.916 (0.143)	0.155 (0.018)			
RE21Y08	0.014	1.127 (0.065)	-0.551 (0.095)	0.204 (0.021)			
RE21Y09	0.019	1.077 (0.050)	-0.633 (0.070)		0.170 (0.041)	-0.170 (0.041)	
RE21Y10	0.031	0.669 (0.053)	0.318 (0.119)				
RE21Y11	0.020	1.238 (0.051)	-0.183 (0.070)	0.193 (0.018)			
RE21Y12	0.024	0.679 (0.033)	-0.236 (0.074)		-0.971 (0.209)	0.971 (0.209)	
RE21Y13	0.021	0.831 (0.041)	0.015 (0.072)		1.081 (0.068)	-0.664 (0.063)	-0.418 (0.063)
RE21Y14	0.021	0.513 (0.034)	-0.049 (0.097)		-0.664 (0.136)	0.664 (0.136)	
RE31M01	0.007	1.173 (0.055)	-1.310 (0.093)	0.180 (0.015)			
RE31M02	0.011	1.126 (0.052)	-1.424 (0.105)				
RE31M03	0.010	0.946 (0.069)	-0.495 (0.106)	0.194 (0.019)			
RE31M04	0.026	0.576 (0.063)	0.063 (0.147)				
RE31M05	0.026	0.985 (0.075)	-0.393 (0.127)	0.347 (0.075)			
RE31M06	0.022	0.833 (0.092)	0.076 (0.176)	0.233 (0.028)			
RE31M07	0.014	1.286 (0.054)	-0.642 (0.117)	0.170 (0.017)			
RE31M08	0.013	1.242 (0.051)	-0.624 (0.080)	0.142 (0.024)			
RE31M09	0.033	0.936 (0.046)	-0.314 (0.063)		1.244 (0.068)	-1.244 (0.068)	

digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for the digitalPIRLS countries, along with the ePIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RE31M10	0.045	0.846 (0.064)	-0.077 (0.106)				
RE31M11	0.016	1.069 (0.055)	-0.948 (0.089)	0.150 (0.020)			
RE31M12	0.025	0.908 (0.063)	0.070 (0.070)	0.133 (0.028)			
RE31M13	0.018	1.457 (0.060)	-1.057 (0.089)	0.182 (0.017)			
RE31M14	0.024	1.602 (0.056)	-0.608 (0.068)	0.126 (0.031)			
RE31M15	0.017	1.122 (0.071)	-0.532 (0.099)	0.188 (0.016)			
RE31M16	0.011	1.091 (0.050)	-0.376 (0.077)				
RE31M17	0.022	0.604 (0.042)	-0.351 (0.077)		0.269 (0.072)	0.129 (0.060)	-0.398 (0.059)
RE31P01	0.013	0.995 (0.069)	-0.639 (0.087)	0.164 (0.017)			
RE31P02	0.025	1.117 (0.072)	-1.317 (0.157)	0.223 (0.025)			
RE31P03	0.026	0.720 (0.040)	-1.062 (0.086)		0.442 (0.070)	-0.442 (0.070)	
RE31P04	0.013	1.297 (0.049)	-0.955 (0.072)				
RE31P05	0.012	1.343 (0.064)	-0.792 (0.076)	0.160 (0.019)			
RE31P06	0.013	0.804 (0.063)	-1.181 (0.182)				
RE31P07	0.025	1.045 (0.051)	-0.143 (0.081)				
RE31P08	0.021	1.305 (0.068)	-0.830 (0.104)	0.177 (0.017)			
RE31P09	0.016	1.289 (0.058)	-0.382 (0.079)	0.150 (0.022)			
RE31P10	0.016	1.136 (0.053)	-0.673 (0.079)	0.190 (0.016)			
RE31P11	0.014	0.798 (0.043)	-0.200 (0.081)		-0.822 (0.101)	0.822 (0.101)	
RE31P12	0.028	0.920 (0.062)	-0.187 (0.084)				
RE31P13	0.022	1.382 (0.065)	-0.656 (0.069)				
RE31P14	0.029	0.911 (0.055)	-0.138 (0.073)				
RE31U01	0.030	0.616 (0.064)	-2.177 (0.203)				
RE31U02	0.008	0.928 (0.049)	-1.969 (0.159)	0.251 (0.034)			
RE31U03	0.014	0.911 (0.073)	-0.869 (0.187)	0.224 (0.030)			
RE31U04	0.024	0.731 (0.053)	-0.966 (0.120)				
RE31U05	0.014	1.224 (0.059)	-1.005 (0.071)	0.147 (0.023)			
RE31U06	0.011	1.246 (0.048)	-0.832 (0.085)	0.156 (0.019)			
RE31U07	0.022	0.980 (0.061)	-0.228 (0.088)	0.127 (0.030)			
RE31U08	0.018	1.194 (0.052)	-0.216 (0.066)	0.120 (0.033)			
RE31U09	0.013	1.271 (0.052)	-1.144 (0.085)	0.164 (0.015)			
RE31U10	0.024	0.998 (0.057)	0.069 (0.085)				
RE31U11	0.033	0.819 (0.057)	-0.960 (0.072)		0.206 (0.063)	-0.206 (0.063)	
RE31U12	0.010	1.124 (0.047)	-1.345 (0.082)				
RE31U13	0.018	1.229 (0.053)	-1.698 (0.104)	0.193 (0.013)			

digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for the digitalPIRLS countries, along with the ePIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RE31U14	0.048	0.525 (0.050)	0.149 (0.127)				
RE31W01	0.037	0.578 (0.040)	-1.031 (0.100)		0.155 (0.075)	-0.155 (0.075)	
RE31W02	0.015	0.878 (0.053)	0.158 (0.062)		0.070 (0.043)	-0.070 (0.043)	
RE31W03	0.008	1.324 (0.074)	0.120 (0.069)	0.098 (0.044)			
RE31W04	0.010	1.218 (0.058)	-0.514 (0.067)				
RE31W05	0.012	1.129 (0.056)	0.138 (0.083)	0.142 (0.024)			
RE31W06	0.014	0.900 (0.062)	-0.895 (0.096)	0.192 (0.016)			
RE31W07	0.021	0.983 (0.043)	0.478 (0.071)		-0.114 (0.056)	0.165 (0.054)	-0.051 (0.046)
RE31W08	0.019	1.121 (0.054)	-0.269 (0.078)	0.207 (0.021)			
RE31W09	0.026	1.035 (0.095)	0.387 (0.103)	0.142 (0.024)			
RE31W10	0.019	1.155 (0.062)	0.249 (0.081)	0.130 (0.029)			
RE31W11	0.021	1.279 (0.056)	0.341 (0.074)				
RE31W12	0.030	1.159 (0.053)	0.458 (0.083)	0.159 (0.017)			
RE31W13	0.023	0.864 (0.045)	0.564 (0.096)				
RE41B01	0.013	0.790 (0.055)	-2.076 (0.143)				
RE41B02	0.014	1.012 (0.065)	-1.563 (0.156)	0.203 (0.018)			
RE41B03	0.029	0.799 (0.046)	-0.580 (0.078)	0.140 (0.024)			
RE41B04	0.017	1.010 (0.054)	-0.675 (0.085)				
RE41B05	0.014	1.122 (0.046)	-0.877 (0.084)	0.145 (0.024)			
RE41B06	0.025	0.800 (0.068)	-0.785 (0.159)	0.146 (0.022)			
RE41B07	0.021	0.922 (0.056)	-0.097 (0.088)				
RE41B08	0.016	0.924 (0.053)	-1.429 (0.093)				
RE41B09	0.014	1.058 (0.057)	-0.347 (0.076)				
RE41B10	0.019	0.682 (0.070)	-0.703 (0.118)	0.189 (0.015)			
RE41B11	0.020	1.091 (0.059)	-0.927 (0.081)	0.152 (0.020)			
RE41B12	0.009	0.929 (0.070)	-1.462 (0.141)	0.241 (0.033)			
RE41B13	0.037	0.687 (0.043)	-0.552 (0.101)		0.534 (0.053)	-0.534 (0.053)	
RE41B14	0.012	1.179 (0.061)	-0.483 (0.072)	0.139 (0.026)			
RE41B15	0.036	0.421 (0.048)	0.316 (0.170)				
RE41B16	0.031	0.478 (0.042)	0.132 (0.092)		0.127 (0.082)	-0.127 (0.082)	
RE41B17	0.022	0.778 (0.059)	-0.311 (0.087)		0.894 (0.077)	-0.894 (0.077)	
RE41E01	0.017	0.833 (0.064)	-1.771 (0.132)				
RE41E02	0.022	0.897 (0.046)	-0.876 (0.136)				
RE41E03	0.020	1.190 (0.052)	-0.737 (0.083)	0.218 (0.023)			
RE41E04	0.016	1.030 (0.050)	-0.372 (0.088)	0.199 (0.021)			

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

digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for the digitalPIRLS countries, along with the ePIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RE41E05	0.010	1.335 (0.056)	-0.954 (0.069)	0.172 (0.018)			
RE41E06	0.014	1.359 (0.050)	-0.954 (0.067)	0.156 (0.020)			
RE41E07	0.033	0.738 (0.043)	-0.222 (0.070)		-0.143 (0.067)	0.143 (0.067)	
RE41E08	0.018	0.860 (0.054)	0.212 (0.096)	0.164 (0.018)			
RE41E09	0.032	0.632 (0.094)	0.067 (0.191)	0.243 (0.038)			
RE41E10	0.023	1.127 (0.059)	-1.342 (0.096)				
RE41E11	0.010	1.290 (0.065)	-0.758 (0.096)	0.163 (0.019)			
RE41E12	0.014	1.278 (0.045)	-1.469 (0.082)				
RE41E13	0.048	0.584 (0.048)	0.195 (0.089)				
RE41E14	0.036	0.463 (0.048)	0.567 (0.127)		1.033 (0.114)	-1.033 (0.114)	
RE41E15	0.022	1.029 (0.055)	-1.348 (0.081)				
RE41E16	0.016	0.888 (0.046)	-0.576 (0.067)				
RE41E17	0.016	1.170 (0.061)	-0.340 (0.097)	0.138 (0.026)			
RE41H01	0.010	0.664 (0.045)	-2.585 (0.177)				
RE41H02	0.013	0.915 (0.058)	-1.576 (0.118)	0.200 (0.013)			
RE41H03	0.026	0.830 (0.053)	-0.123 (0.123)	0.172 (0.018)			
RE41H04	0.014	0.854 (0.077)	-1.625 (0.539)	0.199 (0.013)			
RE41H05	0.018	1.165 (0.047)	-1.908 (0.121)	0.215 (0.019)			
RE41H06	0.013	1.090 (0.052)	-1.047 (0.095)				
RE41H07	0.007	1.367 (0.054)	-0.582 (0.066)	0.133 (0.028)			
RE41H08	0.010	1.098 (0.063)	-0.662 (0.087)	0.201 (0.020)			
RE41H09	0.015	0.872 (0.071)	-1.206 (0.254)	0.188 (0.012)			
RE41H10	0.010	1.031 (0.067)	-0.464 (0.081)	0.197 (0.019)			
RE41H11	0.015	1.005 (0.044)	-0.470 (0.072)				
RE41H12	0.028	0.893 (0.055)	-1.305 (0.134)				
RE41H13	0.015	1.292 (0.066)	-0.509 (0.103)	0.163 (0.020)			
RE41H14	0.011	1.129 (0.074)	-0.785 (0.088)				
RE41H15	0.021	1.102 (0.062)	-0.193 (0.089)	0.172 (0.016)			
RE41H16	0.008	1.122 (0.049)	-1.093 (0.071)	0.162 (0.016)			
RE41I01	0.010	0.748 (0.054)	-2.492 (0.187)				
RE41I02	0.020	0.922 (0.059)	0.347 (0.121)	0.111 (0.037)			
RE41I03	0.024	0.662 (0.037)	0.005 (0.077)		0.340 (0.053)	-0.340 (0.053)	
RE41I04	0.029	0.663 (0.035)	-0.029 (0.078)		0.426 (0.055)	-0.426 (0.055)	
RE41I05	0.014	1.099 (0.043)	-0.143 (0.092)	0.141 (0.024)			
RE41I06	0.019	1.071 (0.094)	0.115 (0.113)	0.187 (0.014)			

digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for the digitalPIRLS countries, along with the ePIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RE41I07	0.020	1.044 (0.050)	0.188 (0.062)		0.049 (0.039)	-0.049 (0.039)	
RE41I08	0.014	1.385 (0.060)	-0.771 (0.074)	0.156 (0.020)			
RE41I09	0.017	1.168 (0.055)	-0.190 (0.069)				
RE41I10	0.014	0.838 (0.059)	-0.137 (0.104)	0.180 (0.017)			
RE41I11	0.020	0.872 (0.059)	0.395 (0.073)		0.496 (0.053)	-0.496 (0.053)	
RE41I12	0.015	1.173 (0.072)	0.116 (0.062)	0.107 (0.039)			
RE41I13	0.019	1.044 (0.052)	-0.126 (0.092)				
RE41I14	0.032	0.836 (0.058)	0.666 (0.114)				
RE41I15	0.013	0.938 (0.056)	0.151 (0.077)				
RE41M01	0.024	0.765 (0.124)	-2.482 (0.594)	0.250 (0.031)			
RE41M02	0.017	0.923 (0.068)	-0.265 (0.106)	0.212 (0.023)			
RE41M03	0.022	0.781 (0.051)	-0.946 (0.088)				
RE41M04	0.018	0.809 (0.078)	-1.260 (0.102)				
RE41M05	0.016	1.178 (0.047)	-1.581 (0.103)	0.213 (0.023)			
RE41M06	0.036	1.070 (0.050)	-1.483 (0.076)				
RE41M07	0.013	0.914 (0.050)	-1.110 (0.116)	0.220 (0.027)			
RE41M08	0.029	1.208 (0.058)	-1.591 (0.091)				
RE41M09	0.020	1.026 (0.058)	-1.007 (0.132)	0.177 (0.015)			
RE41M10	0.023	0.819 (0.048)	-1.576 (0.112)		0.431 (0.071)	-0.431 (0.071)	
RE41M11	0.013	1.099 (0.063)	-0.193 (0.118)	0.174 (0.017)			
RE41M12	0.016	1.005 (0.053)	-1.219 (0.100)				
RE41M13	0.037	1.424 (0.054)	-1.349 (0.078)	0.123 (0.031)			
RE41M14	0.024	0.698 (0.041)	-0.656 (0.084)		0.579 (0.073)	-0.005 (0.074)	-0.574 (0.052)
RE41M15	0.041	0.987 (0.065)	-0.297 (0.107)	0.350 (0.075)			
RE41M16	0.022	1.097 (0.050)	-1.279 (0.085)				
RE41M17	0.006	1.149 (0.070)	-0.537 (0.101)				
RE41M18	0.009	1.031 (0.077)	-1.014 (0.126)	0.192 (0.017)			
RE41O01	0.017	0.951 (0.058)	-1.009 (0.124)	0.201 (0.018)			
RE41O02	0.014	0.927 (0.062)	-1.044 (0.094)				
RE41O03	0.023	0.885 (0.063)	0.685 (0.115)				
RE41O04	0.046	0.598 (0.054)	0.574 (0.086)		-0.151 (0.072)	0.151 (0.072)	
RE41O05	0.021	0.668 (0.049)	-0.054 (0.104)		-0.225 (0.090)	0.225 (0.090)	
RE41O06	0.013	1.313 (0.047)	-0.507 (0.089)	0.170 (0.017)			
RE41O07	0.025	0.784 (0.048)	-0.845 (0.091)		0.497 (0.056)	-0.497 (0.056)	
RE41O08	0.022	0.896 (0.047)	-0.072 (0.087)				

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

digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for the digitalPIRLS countries, along with the ePIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RE41O09	0.027	1.209 (0.054)	-0.246 (0.064)				
RE41O10	0.019	0.926 (0.049)	0.098 (0.063)		0.194 (0.040)	-0.194 (0.040)	
RE41O11	0.012	1.198 (0.054)	0.052 (0.073)	0.131 (0.029)			
RE41O12	0.015	1.345 (0.059)	-0.262 (0.064)	0.098 (0.044)			
RE41O13	0.019	0.643 (0.045)	0.327 (0.060)		-1.183 (0.257)	0.720 (0.187)	0.463 (0.116)
RE51C01	0.025	0.496 (0.038)	0.476 (0.093)		-0.090 (0.100)	0.090 (0.100)	
RE51C02	0.026	0.824 (0.054)	-1.073 (0.143)	0.222 (0.025)			
RE51C03	0.022	1.031 (0.091)	0.552 (0.146)				
RE51C04	0.011	0.979 (0.084)	-1.196 (0.161)	0.170 (0.013)			
RE51C05	0.014	0.873 (0.046)	-0.646 (0.072)		0.190 (0.069)	-0.190 (0.069)	
RE51C06	0.026	0.665 (0.057)	0.321 (0.073)		-0.100 (0.085)	0.100 (0.085)	
RE51C07	0.021	0.838 (0.047)	-0.615 (0.088)				
RE51C08	0.013	1.090 (0.062)	-0.217 (0.078)	0.153 (0.021)			
RE51C09	0.011	1.356 (0.060)	-0.797 (0.076)	0.151 (0.021)			
RE51C10	0.020	0.736 (0.061)	-0.447 (0.086)		0.626 (0.072)	-0.626 (0.072)	
RE51C11	0.015	1.031 (0.049)	-0.059 (0.068)	0.144 (0.024)			
RE51C12	0.012	1.284 (0.071)	-0.393 (0.118)	0.176 (0.016)			
RE51C13	0.019	0.935 (0.069)	-0.024 (0.082)		0.328 (0.054)	-0.328 (0.054)	
RE51C14	0.030	0.953 (0.049)	0.873 (0.058)		0.088 (0.057)	-0.088 (0.057)	
RE51C15	0.025	0.909 (0.119)	0.994 (0.222)				
RE51D01	0.021	0.710 (0.061)	-1.500 (0.121)				
RE51D02	0.025	0.764 (0.072)	0.280 (0.099)	0.217 (0.023)			
RE51D03	0.017	0.858 (0.058)	-0.405 (0.102)				
RE51D04	0.005	1.270 (0.054)	-0.449 (0.090)	0.142 (0.024)			
RE51D05	0.018	0.943 (0.061)	-0.594 (0.073)	0.138 (0.025)			
RE51D06	0.051	0.701 (0.037)	-0.294 (0.066)		0.593 (0.065)	-0.593 (0.065)	
RE51D07	0.010	1.119 (0.095)	0.249 (0.091)				
RE51D08	0.020	1.130 (0.087)	-0.274 (0.074)	0.130 (0.029)			
RE51D09	0.016	1.073 (0.065)	0.186 (0.128)	0.155 (0.020)			
RE51D10	0.017	0.560 (0.041)	0.185 (0.074)		-1.704 (0.233)	1.704 (0.233)	
RE51D11	0.036	0.706 (0.047)	-0.044 (0.072)		0.479 (0.051)	-0.479 (0.051)	
RE51D12	0.013	0.694 (0.037)	-0.357 (0.067)		0.638 (0.077)	0.231 (0.070)	-0.869 (0.081)
RE51D13	0.022	0.858 (0.043)	-0.417 (0.073)				
RE51D14	0.014	1.045 (0.046)	-0.090 (0.099)	0.162 (0.018)			
RE51D15	0.030	0.601 (0.067)	0.920 (0.116)		-0.224 (0.204)	0.224 (0.204)	

digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for the digitalPIRLS countries, along with the ePIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RE51D16	0.023	0.941 (0.068)	0.010 (0.103)				
RE51N01	0.013	0.853 (0.055)	-1.333 (0.103)				
RE51N02	0.022	0.637 (0.035)	-0.017 (0.075)		0.048 (0.061)	-0.048 (0.061)	
RE51N03	0.027	0.946 (0.072)	0.785 (0.099)	0.148 (0.021)			
RE51N04	0.028	0.669 (0.048)	-0.729 (0.118)				
RE51N05	0.019	0.865 (0.047)	0.379 (0.109)				
RE51N06	0.016	0.761 (0.040)	-0.091 (0.061)		-0.037 (0.042)	0.037 (0.042)	
RE51N07	0.022	0.602 (0.047)	-0.748 (0.163)	0.195 (0.015)			
RE51N08	0.012	1.023 (0.053)	-0.449 (0.098)	0.178 (0.019)			
RE51N09	0.020	0.881 (0.053)	0.377 (0.076)		0.062 (0.061)	-0.062 (0.061)	
RE51N10	0.020	0.905 (0.062)	0.954 (0.079)				
RE51N11	0.012	1.011 (0.065)	0.167 (0.100)				
RE51N12	0.021	0.704 (0.041)	0.598 (0.099)				
RE51N13	0.015	0.879 (0.050)	-0.386 (0.082)		0.230 (0.033)	-0.230 (0.033)	
RE51N14	0.016	0.988 (0.073)	0.543 (0.080)				
RE51N15	0.032	0.812 (0.085)	1.003 (0.145)				
RE51R01	0.019	0.685 (0.068)	-1.505 (0.235)	0.205 (0.014)			
RE51R02	0.016	0.792 (0.050)	-2.106 (0.134)				
RE51R03	0.011	0.773 (0.038)	-3.612 (0.192)				
RE51R04	0.021	0.774 (0.056)	-0.373 (0.124)				
RE51R05	0.017	0.747 (0.044)	-0.751 (0.079)		-0.323 (0.065)	0.323 (0.065)	
RE51R06	0.017	1.072 (0.062)	-0.843 (0.122)	0.213 (0.026)			
RE51R07	0.017	0.945 (0.052)	-1.087 (0.097)				
RE51R08	0.010	1.021 (0.055)	-0.758 (0.089)	0.221 (0.027)			
RE51R09	0.019	0.967 (0.048)	-0.227 (0.105)	0.162 (0.018)			
RE51R10	0.016	1.160 (0.050)	0.130 (0.102)	0.102 (0.042)			
RE51R11	0.021	0.911 (0.066)	-0.542 (0.118)	0.235 (0.032)			
RE51R13	0.027	0.621 (0.068)	-0.089 (0.118)				
RE51R14	0.026	0.834 (0.077)	-0.804 (0.145)	0.296 (0.063)			
RE51R15	0.008	1.193 (0.045)	-0.629 (0.076)				
RE51R16	0.035	0.616 (0.051)	-0.569 (0.138)				
RE51R17	0.026	0.732 (0.043)	-0.483 (0.088)		0.586 (0.058)	-0.586 (0.058)	
RE51T01	0.017	0.736 (0.067)	-2.587 (0.193)				
RE51T02	0.024	0.755 (0.048)	0.515 (0.114)	0.179 (0.012)			
RE51T03	0.016	0.711 (0.069)	-0.669 (0.105)	0.190 (0.016)			

digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for the digitalPIRLS countries, along with the ePIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
RE51T04	0.020	0.817 (0.053)	-0.757 (0.095)				
RE51T05	0.023	0.675 (0.038)	-0.232 (0.070)		-0.224 (0.056)	0.224 (0.056)	
RE51T06	0.009	1.319 (0.051)	-0.612 (0.071)	0.141 (0.026)			
RE51T07	0.010	1.255 (0.044)	-0.556 (0.062)	0.160 (0.019)			
RE51T08	0.007	1.135 (0.051)	-1.121 (0.089)	0.160 (0.018)			
RE51T09	0.022	0.867 (0.070)	-0.394 (0.141)	0.203 (0.023)			
RE51T10	0.012	1.174 (0.055)	-0.389 (0.077)				
RE51T11	0.021	0.862 (0.054)	-0.658 (0.065)		-0.617 (0.088)	0.617 (0.088)	
RE51T12	0.024	0.787 (0.081)	-0.844 (0.119)				
RE51T13	0.042	0.796 (0.051)	0.231 (0.070)		0.728 (0.075)	-0.728 (0.075)	
RE51T14	0.017	0.984 (0.040)	-0.015 (0.073)				
RE51T15	0.027	0.678 (0.037)	0.097 (0.078)		0.605 (0.071)	-0.605 (0.071)	
RE51T16	0.027	0.799 (0.059)	0.203 (0.091)		0.540 (0.076)	-0.540 (0.076)	
RE51Z01	0.025	0.509 (0.038)	-1.108 (0.110)		0.180 (0.094)	-0.180 (0.094)	
RE51Z02	0.030	0.559 (0.047)	-1.899 (0.195)				
RE51Z03	0.013	1.078 (0.053)	-0.238 (0.076)	0.129 (0.029)			
RE51Z04	0.014	1.021 (0.052)	-0.567 (0.078)	0.169 (0.020)			
RE51Z05	0.014	1.203 (0.059)	0.162 (0.063)				
RE51Z06	0.014	0.667 (0.041)	-1.647 (0.095)		0.047 (0.091)	-0.047 (0.091)	
RE51Z07	0.019	1.454 (0.051)	-0.945 (0.064)				
RE51Z08	0.023	1.472 (0.057)	-0.494 (0.056)	0.113 (0.037)			
RE51Z09	0.043	0.958 (0.049)	0.280 (0.092)				
RE51Z10	0.013	1.218 (0.053)	-0.611 (0.070)				
RE51Z11	0.016	1.321 (0.058)	-0.958 (0.081)				
RE51Z12	0.013	0.941 (0.051)	-0.488 (0.072)		-0.009 (0.056)	0.009 (0.056)	
RE51Z13	0.016	1.175 (0.057)	-0.817 (0.091)	0.167 (0.017)			
RE51Z14	0.020	0.558 (0.042)	-0.688 (0.070)		-0.160 (0.129)	-0.016 (0.116)	0.176 (0.103)
RE51Z15	0.021	0.589 (0.059)	-0.462 (0.135)				
Non-Response Indicators							
RECRLIT	0.011	1.626 (0.208)	0.076 (0.078)	0.602 (0.030)			
RECRINF	0.013	1.839 (0.254)	0.381 (0.055)	0.584 (0.024)			
REMCLIT	0.003	1.989 (0.376)	-0.165 (0.143)	0.892 (0.015)			
REMCINF	0.003	2.259 (0.609)	0.151 (0.123)	0.898 (0.012)			
RECRRSI	0.011	2.045 (0.282)	-0.045 (0.081)	0.620 (0.031)			
RECRIIE	0.012	1.636 (0.229)	0.364 (0.060)	0.528 (0.026)			

digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for the digitalPIRLS countries, along with the ePIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
REMCRSI	0.004	2.148 (0.483)	-0.028 (0.126)	0.870 (0.015)			
REMCIIE	0.004	2.189 (0.644)	-0.035 (0.166)	0.893 (0.017)			

( ) Standard errors appear in parentheses.

ePIRLS item parameters estimated from PIRLS 2021 ePIRLS calibration using fixed digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for digitalPIRLS countries, along with digitalPIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
				ePIRLS Items			
E041R01	0.006	0.842 (0.075)	0.177 (0.126)	0.187 (0.023)			
E041R02	0.016	0.612 (0.051)	-1.509 (0.147)				
E041R03	0.020	0.558 (0.036)	-0.490 (0.100)		0.343 (0.070)	-0.343 (0.070)	
E041R04	0.013	1.551 (0.176)	0.755 (0.079)	0.152 (0.015)			
E041R05	0.020	0.504 (0.053)	-0.785 (0.144)				
E041R06	0.015	0.675 (0.044)	-1.246 (0.093)				
E041R07	0.020	0.939 (0.054)	0.057 (0.034)		-0.013 (0.027)	0.013 (0.027)	
E041R08	0.009	0.743 (0.048)	0.330 (0.074)				
E041R09	0.011	0.600 (0.045)	-0.171 (0.069)				
E041R10	0.005	1.005 (0.068)	-0.371 (0.090)	0.222 (0.035)			
E041R11	0.010	0.690 (0.043)	0.225 (0.072)		0.259 (0.055)	-0.259 (0.055)	
E041R12	0.006	1.555 (0.138)	-0.239 (0.055)	0.249 (0.027)			
E041R13	0.006	0.790 (0.078)	0.102 (0.090)	0.211 (0.030)			
E041R14	0.006	0.820 (0.044)	-0.275 (0.080)				
E041R15	0.012	1.302 (0.059)	-0.136 (0.031)				
E041R16	0.017	0.804 (0.050)	-0.174 (0.069)				
E041T01	0.006	0.898 (0.178)	0.258 (0.117)	0.270 (0.047)			
E041T02	0.016	0.960 (0.061)	-1.635 (0.107)				
E041T03	0.015	0.681 (0.127)	0.167 (0.096)	0.159 (0.023)			
E041T04	0.014	1.195 (0.088)	-0.875 (0.076)	0.188 (0.025)			
E041T05	0.009	0.898 (0.042)	-0.836 (0.045)		0.252 (0.047)	-0.252 (0.047)	
E041T06	0.011	0.953 (0.072)	-0.245 (0.064)				
E041T08	0.007	0.753 (0.065)	-0.306 (0.071)				
E041T09	0.012	0.976 (0.057)	0.749 (0.068)				
E041T10	0.011	1.103 (0.060)	-0.060 (0.088)				
E041T11	0.018	1.009 (0.061)	-0.612 (0.058)	0.148 (0.019)			
E041T12	0.010	0.608 (0.054)	-1.066 (0.169)	0.214 (0.027)			
E041T13	0.007	1.095 (0.114)	0.057 (0.105)	0.276 (0.039)			
E041T14	0.006	0.676 (0.071)	0.272 (0.080)				
E041T15	0.014	1.834 (0.112)	-0.008 (0.038)	0.120 (0.012)			
E041T16	0.010	1.172 (0.072)	-0.734 (0.077)	0.188 (0.034)			
E041T17	0.010	1.181 (0.090)	-0.163 (0.075)				
E041T18	0.047	0.746 (0.071)	0.414 (0.066)				

ePIRLS item parameters estimated from PIRLS 2021 ePIRLS calibration using fixed digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for digitalPIRLS countries, along with digitalPIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
E041Z01	0.005	0.987 (0.122)	-0.196 (0.133)	0.296 (0.054)			
E041Z02	0.012	0.584 (0.038)	-0.188 (0.085)		-0.081 (0.091)	0.081 (0.091)	
E041Z03	0.009	0.800 (0.072)	0.071 (0.119)	0.197 (0.036)			
E041Z04	0.018	0.906 (0.081)	-1.141 (0.088)				
E041Z05	0.017	0.560 (0.070)	-0.422 (0.142)	0.175 (0.015)			
E041Z06	0.007	1.191 (0.074)	-0.181 (0.042)				
E041Z07	0.011	1.388 (0.120)	-0.078 (0.069)	0.262 (0.033)			
E041Z08	0.005	1.163 (0.137)	-0.009 (0.078)	0.270 (0.037)			
E041Z09	0.007	0.614 (0.048)	0.638 (0.134)				
E041Z10	0.010	1.112 (0.100)	0.012 (0.086)	0.176 (0.023)			
E041Z11	0.010	1.382 (0.099)	-0.807 (0.081)	0.244 (0.037)			
E041Z12	0.010	0.737 (0.047)	-0.585 (0.082)				
E041Z13	0.012	1.416 (0.094)	-0.630 (0.076)	0.184 (0.024)			
E041Z14	0.020	0.811 (0.066)	-0.780 (0.082)		0.326 (0.064)	-0.326 (0.064)	
E041Z15	0.010	1.300 (0.141)	-0.083 (0.075)	0.270 (0.033)			
E041Z16	0.022	0.819 (0.044)	-0.289 (0.051)		0.227 (0.040)	-0.227 (0.040)	
E041Z17	0.011	0.700 (0.065)	-0.005 (0.104)				
E041Z18	0.006	1.149 (0.096)	0.086 (0.078)	0.181 (0.025)			
E041Z19	0.018	1.015 (0.045)	0.349 (0.043)		0.406 (0.032)	-0.406 (0.032)	
E041Z20	0.014	0.893 (0.061)	-0.240 (0.048)		0.331 (0.035)	-0.331 (0.035)	
E051O01	0.009	1.072 (0.087)	0.010 (0.073)	0.178 (0.026)			
E051O02	0.005	1.225 (0.162)	-0.679 (0.210)	0.383 (0.085)			
E051O03	0.008	1.128 (0.269)	-0.051 (0.123)	0.347 (0.060)			
E051O04	0.011	0.698 (0.073)	-0.559 (0.120)	0.176 (0.021)			
E051O05	0.007	1.086 (0.089)	-0.515 (0.077)				
E051O06	0.009	1.035 (0.096)	-0.580 (0.133)	0.244 (0.057)			
E051O07	0.007	0.924 (0.108)	-0.099 (0.118)	0.312 (0.039)			
E051O08	0.018	0.439 (0.036)	-1.053 (0.110)		-1.119 (0.159)	1.119 (0.159)	
E051O09	0.012	1.288 (0.070)	-0.766 (0.045)				
E051O10	0.021	0.633 (0.040)	-0.716 (0.060)		0.259 (0.079)	-0.259 (0.079)	
E051O11	0.010	0.517 (0.049)	0.273 (0.093)				
E051O12	0.022	0.451 (0.030)	0.333 (0.099)		0.681 (0.100)	-0.681 (0.100)	
E051O13	0.016	0.613 (0.052)	-1.018 (0.068)		0.153 (0.055)	-0.153 (0.055)	
E051O14	0.013	0.992 (0.153)	-0.539 (0.124)	0.266 (0.057)			

ePIRLS item parameters estimated from PIRLS 2021 ePIRLS calibration using fixed digitalPIRLS item parameters from the PIRLS 2021 digital calibration. Used to estimate student proficiency for digitalPIRLS countries, along with digitalPIRLS items.

Item	RMSD	Slope (a_i)	Location (b_i)	Guessing (c_i)	Step 1 (t_{i1})	Step 2 (t_{i2})	Step 3 (t_{i3})
E051O15	0.013	1.457 (0.165)	1.001 (0.059)	0.082 (0.010)			
E051O16	0.008	1.588 (0.171)	0.374 (0.068)	0.241 (0.024)			
E051O17	0.007	1.104 (0.073)	0.477 (0.050)				
E051O18	0.020	0.341 (0.034)	0.672 (0.110)		-0.108 (0.090)	0.108 (0.090)	
E051V01	0.010	0.703 (0.150)	0.424 (0.125)	0.213 (0.046)			
E051V02	0.020	1.056 (0.118)	-0.426 (0.105)	0.192 (0.029)			
E051V03	0.012	0.789 (0.090)	0.430 (0.090)				
E051V04	0.012	0.449 (0.034)	0.932 (0.102)		0.542 (0.187)	-0.542 (0.187)	
E051V05	0.005	1.143 (0.146)	-0.187 (0.111)	0.349 (0.051)			
E051V06	0.006	0.983 (0.055)	0.067 (0.051)				
E051V07	0.007	1.145 (0.113)	-0.223 (0.101)	0.202 (0.038)			
E051V08	0.010	1.953 (0.137)	0.147 (0.056)	0.225 (0.021)			
E051V09	0.017	0.667 (0.036)	-0.093 (0.044)		0.161 (0.049)	-0.161 (0.049)	
E051V10	0.007	0.812 (0.066)	0.814 (0.104)				
E051V11	0.013	1.311 (0.087)	-0.185 (0.044)	0.141 (0.018)			
E051V12	0.008	0.547 (0.145)	0.618 (0.149)	0.175 (0.035)			
E051V13	0.018	1.466 (0.081)	-1.590 (0.072)				
E051V14	0.016	0.992 (0.102)	-0.034 (0.088)	0.226 (0.035)			
E051V15	0.016	0.720 (0.045)	0.323 (0.083)		-0.062 (0.041)	0.062 (0.041)	
E051V16	0.012	1.239 (0.075)	-0.013 (0.056)				
E051V17	0.006	0.598 (0.041)	0.683 (0.054)		0.365 (0.042)	-0.365 (0.042)	
E051V18	0.006	1.036 (0.083)	0.431 (0.054)				
E051V19A	0.015	0.708 (0.054)	1.736 (0.134)				
E051V20	0.010	0.910 (0.073)	1.024 (0.073)				
Non-Response Indicators							
E0CRPT1	0.005	1.620 (0.254)	0.373 (0.053)	0.462 (0.035)			
E0CRPT2	0.007	1.838 (0.211)	0.370 (0.049)	0.539 (0.027)			
E0MCPT1	0.005	1.379 (0.326)	0.052 (0.161)	0.747 (0.032)			
E0MCPT2	0.005	0.667 (0.136)	-0.734 (0.482)	0.653 (0.087)			
E0CRRSI	0.004	1.643 (0.219)	-0.102 (0.081)	0.586 (0.036)			
E0CRIIE	0.008	1.618 (0.178)	0.441 (0.046)	0.435 (0.029)			
E0MCRSI	0.007	0.804 (0.147)	-0.296 (0.264)	0.581 (0.061)			
E0MCIIE	0.002	2.787 (0.689)	0.026 (0.092)	0.880 (0.013)			

( ) Standard errors appear in parentheses.