CHAPTER 12

Scaling the PIRLS 2016 Achievement Data

Pierre Foy
Liqun Yin

Overview

The PIRLS 2016 assessment had ambitious goals for broad coverage of the reading purposes and processes, as described in its assessment framework, and for measuring trends across assessment cycles. Given this broad coverage, PIRLS used a matrix-sampling booklet design such that each student was administered only a subset of the entire PIRLS item pool (see Chapter 4 of *PIRLS 2016 Assessment Framework, 2nd Edition*). Given the complexities of the data collection and the need to have student scores on the entirety of each assessment for analysis and reporting purposes, PIRLS relied on item response theory (IRT) scaling to describe student achievement and to provide accurate measures of trends. As each student responded to only a part of the assessment item pool, the PIRLS scaling approach used multiple imputation—or plausible values—methodology to obtain proficiency scores in reading for all students. To enhance the reliability of the student scores, the PIRLS scaling approach uses conditioning, a process in which student responses to the items are combined with information about students' backgrounds.

This scaling chapter begins with a general description of the PIRLS scaling approach and its use of plausible values. It then describes the concurrent calibration method used specifically to measure trends. Next, it explains how the proficiency scores are generated through the use of conditioning and describes the process of transforming the proficiency scores to place them on the metric used to measure trends. A special section then describe how the PIRLS Literacy 2016 achievement data were scaled and placed on the PIRLS reading reporting scale and another section describes the scaling of the ePIRLS 2016 achievement data. A theoretical description of the PIRLS scaling methodology can be found in Chapter 11: PIRLS 2016 Achievement Scaling Methodology.

Implementing the PIRLS Scaling Procedure

The application of IRT scaling and plausible values methodology to the data from the PIRLS 2016 assessment involved four major tasks: calibrating the achievement items (estimating model parameters for each item), creating principal components from the student and parent questionnaire data for use in conditioning, generating reading proficiency scores, and placing these proficiency scores on the metric used to report trend results from previous assessments. New for PIRLS 2016, the PIRLS Literacy achievement results were reported on the PIRLS reading scale. Also, in order to report trends back to its predecessor assessment, prePIRLS 2011, the 2011 scores were re-calibrated. The scaling procedure also generated proficiency scores for the domains of overall reading: the purposes for reading and the processes of comprehension.

Linking Assessments Cycles with Concurrent Calibration

The metric of the PIRLS reporting scale was originally established in PIRLS 2001 by setting the mean of the national average scores for all countries that participated in PIRLS 2001 to 500 and the standard deviation to 100. To enable measurement of trends over time, achievement data from successive PIRLS assessments were transformed to this same metric. This is done by concurrently scaling the data from each successive assessment with the data from the previous assessment—a process known as concurrent calibration—and applying linear transformations to place the results from each successive assessment on the same scale as the results from the previous assessment. This procedure enables PIRLS to measure trends across all four assessment cycles: 2001, 2006, 2011, and 2016.[1]

The first step in linking the assessments for trend scaling is to estimate (calibrate) the item parameters for the items in the current assessment through a concurrent calibration of the data from the current assessment and from the previous assessment. In 2016, the PIRLS concurrent calibration consisted of combining achievement data from the 2016 and 2011 assessments.

In linking successive assessments, concurrent calibration relies on having a large proportion of trend items, items that are retained from one assessment to the next. The PIRLS 2016 assessment consisted of 6 literary passages with their items and 6 informational passages with their items. In PIRLS 2016, 3 of the literary passages and 3 of the informational passages consisted of newly developed items. The remaining 3 literary passages and 3 informational passages were carried forward from the PIRLS 2011 assessment and are the basis for linking PIRLS 2016 to the PIRLS achievement scale and maintaining trends over time. Exhibit 12.1 lists the number of items present for the PIRLS 2016 concurrent calibration by item type and by purposes for reading and processes of comprehension.

1 See Mazzeo and von Davier (2014) for a discussion of the linking procedure used by PIRLS.

Exhibit 12.1: PIRLS 2016 Reading Items for Concurrent Calibration

Item Type	Points	Items Released in 2011		Items Common in 2011 and 2016		Items Introduced in 2016		Total	
		Items	Points	Items	Points	Items	Points	Items	Points
Multiple-Choice	1	29	29	45	45	41	41	115	115
Constructed Response	1	12	12	16	16	33	33	61	61
	2	12	24	15	30	17	34	44	88
	3	1	3	5	15	3	9	9	27
Total		**54**	**68**	**81**	**106**	**94**	**117**	**229**	**291**

PIRLS 2016 Reading Items for Concurrent Calibration by Reading Purposes and Comprehension Processes

Purposes for Reading Items	Items Released in 2011		Items Common in 2011 and 2016		Items Introduced in 2016		Total	
	Items	Points	Items	Points	Items	Points	Items	Points
Literary Experience	28	35	44	55	46	58	118	148
Acquire and Use Information	26	33	37	51	48	59	111	143
Processes of Comprehension	**Items Released in 2011**		**Items Common in 2011 and 2016**		**Items Introduced in 2016**		**Total**	
	Items	Points	Items	Points	Items	Points	Items	Points
Retrieving and Straightforward Inferencing	33	36	46	51	57	65	136	152
Interpreting, Integrating, and Evaluating	21	32	35	55	37	52	93	139
Total	**54**	**68**	**81**	**106**	**94**	**117**	**229**	**291**

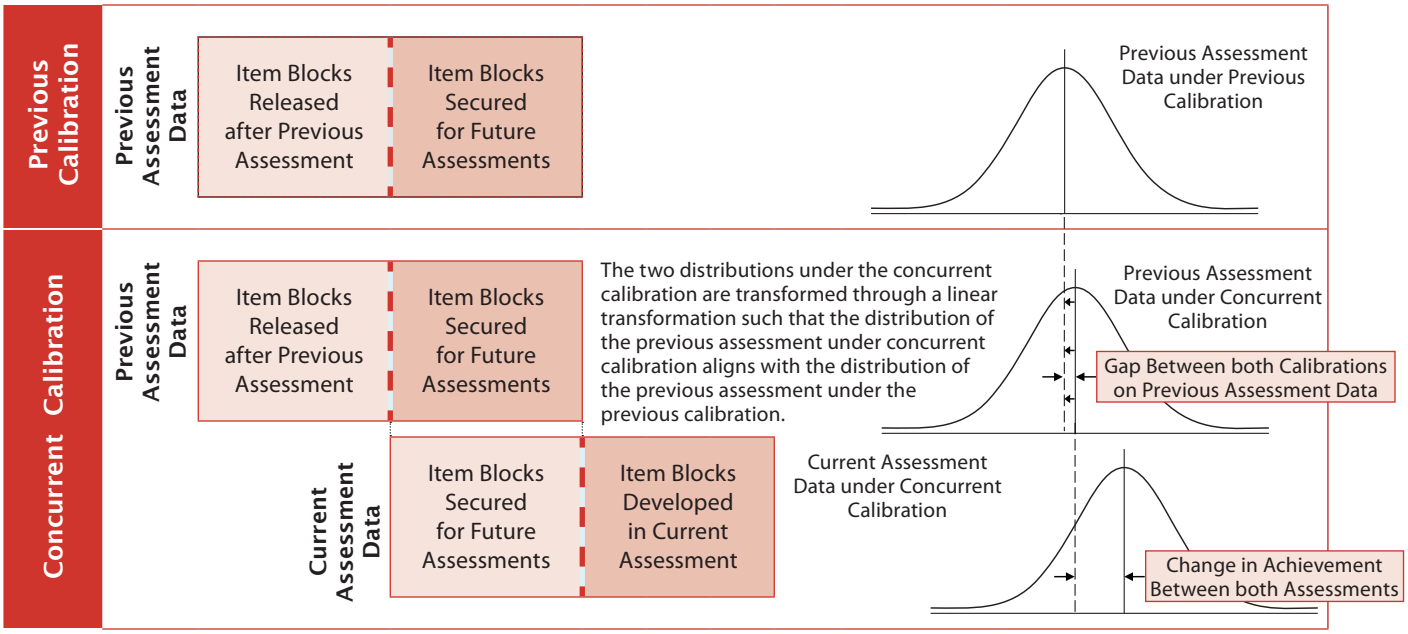
In concurrent calibration, item parameters for the current assessment are estimated based on the data from both the current and previous assessments, recognizing that some items (the trend items) are common to both. It is then possible to estimate the latent ability distributions of students in both assessments using the item parameters from the concurrent calibration. The difference between these two distributions is the change in achievement between the previous and current assessments.

After the concurrent calibration, the next step is to make use of student and parent context data from their respective questionnaires, in a process called conditioning, to enhance the reliability of the proficiency scores estimated. Once these proficiency scores are estimated, the next step consists of finding a linear transformation that transforms the proficiency distribution of the previous assessment data under the concurrent calibration to match the proficiency distribution

of these same data under the calibration that was done in the previous assessment. The final step entails applying this linear transformation to the current assessment data scaled using the concurrent calibration. This places the current assessment data on the trend scale.

Exhibit 12.2 illustrates how the concurrent calibration approach is applied in the context of PIRLS trend scaling. The gap between the distributions of the previous assessment data under the previous calibration and under the concurrent calibration is typically small and is the result of slight differences in the item parameter estimates from the two calibrations (Exhibit 12.2, second panel). The linear transformation removes this gap by shifting the two distributions from the concurrent calibration such that the distribution of the previous assessment data from the concurrent calibration aligns with the distribution of the previous assessment data from the previous calibration,[2] while preserving the gap between the previous and current assessment data under the concurrent calibration. This latter gap is the change in achievement between the previous and current assessments that PIRLS sets out to measure as trend.

Exhibit 12.2: Concurrent Calibration Model Used for PIRLS



Calibrating the PIRLS 2016 Assessment Data

Item calibration was conducted by the TIMSS & PIRLS International Study Center using the commercially-available Parscale software (Muraki & Bock, 1991) and included data from the previous assessment (PIRLS 2011) and data from the 2016 assessment for countries that participated

2 The difference between the ability distributions of the previous assessment data under the two calibrations is a measure of the linkage error in the trend scaling procedure.

in both assessment cycles. The calibration used all available item response data from each country's student samples and from both current and previous assessments. All student samples were weighted so that each country contributed equally to the item calibration. Exhibit 12.3 shows the sample sizes for scaling the PIRLS 2016 data. A total of 40 countries from PIRLS 2016 contributed to the concurrent calibration. Norway's data at the fourth grade were included in the concurrent calibration.

Exhibit 12.3: Sample Sizes for PIRLS 2016 Achievement Scales

Country	Concurrent Calibration		Proficiency Estimation	
	2016	2011	2016	2011
Australia	6,341	6,126	6,341	6,126
Austria	4,360	4,670	4,360	4,670
Azerbaijan	4,990	4,881	5,994	4,881
Bahrain	—	—	5,480	—
Belgium (Flemish)	—	—	5,198	—
Belgium (French)	4,623	3,727	4,623	3,727
Bulgaria	4,281	5,261	4,281	5,261
Canada	18,245	23,206	18,245	23,206
Chile	—	—	4,294	—
Chinese Taipei	4,326	4,293	4,326	4,293
Czech Republic	5,537	4,556	5,537	4,556
Denmark	3,508	4,594	3,508	4,594
England	5,095	3,927	5,095	3,927
Finland	4,896	4,640	4,896	4,640
France	4,767	4,438	4,767	4,438
Georgia	5,741	4,796	5,741	4,796
Germany	3,959	4,000	3,959	4,000
Hong Kong SAR	3,349	3,875	3,349	3,875
Hungary	4,623	5,204	4,623	5,204
Iran, Islamic Rep. of	4,385	5,758	4,385	5,758
Ireland	4,607	4,524	4,607	4,524
Israel	4,041	4,186	4,041	4,186
Italy	3,940	4,189	3,940	4,189
Kazakhstan	—	—	4,925	—
Latvia	—	—	4,157	—
Lithuania	2,947	4,661	4,317	4,661
Macao SAR	—	—	4,059	—

Exhibit 12.3: Sample Sizes for PIRLS 2016 Achievement Scales (Continued)

Country	Concurrent Calibration		Proficiency Estimation	
	2016	2011	2016	2011
Malta	3,647	3,548	3,647	3,548
Morocco	5,489	7,805	5,489	7,805
Netherlands	4,206	3,995	4,206	3,995
New Zealand	5,646	5,644	5,646	5,644
Northern Ireland	3,693	3,586	3,693	3,586
Norway (5)	—	—	4,232	—
Oman	9,234	10,394	9,234	10,394
Poland	—	—	4,413	—
Portugal	4,642	4,085	4,642	4,085
Qatar	9,077	4,120	9,077	4,120
Russian Federation	4,577	4,461	4,577	4,461
Saudi Arabia	4,741	4,507	4,741	4,507
Singapore	6,488	6,367	6,488	6,367
Slovak Republic	5,451	5,630	5,451	5,630
Slovenia	4,499	4,512	4,499	4,512
Spain	14,595	8,580	14,595	8,580
Sweden	4,525	4,622	4,525	4,622
Trinidad and Tobago	4,177	3,948	4,177	3,948
United Arab Emirates	16,471	14,618	16,471	14,618
United States	4,425	12,726	4,425	12,726
Benchmarking Participants				
Buenos Aires, Argentina	—	—	4,382	—
Ontario, Canada	—	—	4,270	4,561
Quebec, Canada	—	—	3,179	4,244
Norway (4)	4,354	3,190	4,354	3,190
Moscow City, Russian Fed.	—	—	4,289	—
Eng/Afr/Zulu - RSA (5)	—	—	5,282	—
Andalusia, Spain	—	—	4,169	4,333
Madrid, Spain	—	—	3,794	—
Abu Dhabi, UAE	—	—	4,188	4,146
Dubai, UAE	—	—	7,859	6,061
Total	**228,498**	**231,850**	**309,042**	**255,195**

The item parameters estimated from these concurrent calibrations, based on the countries that have participated in both the previous and current assessments, were used to estimate student proficiency for all countries and benchmarking entities participating in the PIRLS 2016 assessment. These item parameters also were used to estimate student proficiency in the purposes for reading and processes of comprehension domains. Student proficiency was estimated for a total of 47 countries and 10 benchmarking participants, as shown in Exhibit 12.3. The item parameters estimated from the PIRLS 2016 concurrent calibration are presented in Appendix 13A.

Treatment of Omitted and Not-Reached Responses

Given the matrix-sampling design used by PIRLS, whereby a student is administered only a sample of the assessment items (from one literary passage and one informational passage) most items are missing by design for each student. However, missing data can also result from a student not answering an item, which can occur when the student does not know the answer, omits the item by mistake, or does not have sufficient time to attempt the item. An item is considered "not reached" when—within part 1 or part 2 of a booklet[3]—the item itself and the item immediately preceding it are not answered, and there are no other items completed in the remainder of that part of the booklet.

Not-reached items are treated differently in estimating item parameters and in generating student proficiency scores. In estimating the values of the item parameters, items in the assessment booklets that are considered not to have been reached by students are treated as if they have not been administered. This approach is considered optimal for parameter estimation. However, not-reached items are considered as incorrect responses when student proficiency scores are generated.

Evaluating Fit of IRT Models to the PIRLS Assessment Data

After the item calibration was completed, checks were performed to verify that the item parameters obtained from Parscale adequately reproduced the observed distribution of student responses across the proficiency continuum. The fit of the IRT models to the PIRLS assessment data was examined by comparing the item response function curves generated using the item parameters estimated from the data with the empirical item response functions calculated from the latent abilities estimated for each student that responded to the item. When the empirical results for an item fall near the fitted curves, the IRT model fits the data well and provides an accurate and reliable measurement of the underlying proficiency scale. Graphical plots of these response function curves are called item characteristic curves (ICC).

The plots in the Exhibits 12.4 and 12.5 show examples of the empirical and fitted item response functions for dichotomously scored (right/wrong) multiple-choice and constructed response items,

3 The PIRLS assessment consist of two parts, with a break in between.

respectively. In each plot, the horizontal axis represents the proficiency scale, and the vertical axis represents the probability of a correct response. The fitted curve based on the estimated item parameters is shown as a solid line. Empirical results are represented by circles. The empirical results are obtained by first dividing the proficiency scale into intervals of equal size and then counting the number of students responding to the item whose estimated latent abilities (EAP scores) from Parscale fall in each interval. Then the proportion of students in each interval that responded correctly to the item is calculated. In the exhibits, the center of each circle represents this empirical proportion of correct responses. The size of each circle is proportional to the number of students contributing to the estimation of the empirical proportion correct.

Exhibit 12.4: Example of Item Response Function for a Dichotomous Multiple-Choice Item from the PIRLS 2016 Assessment

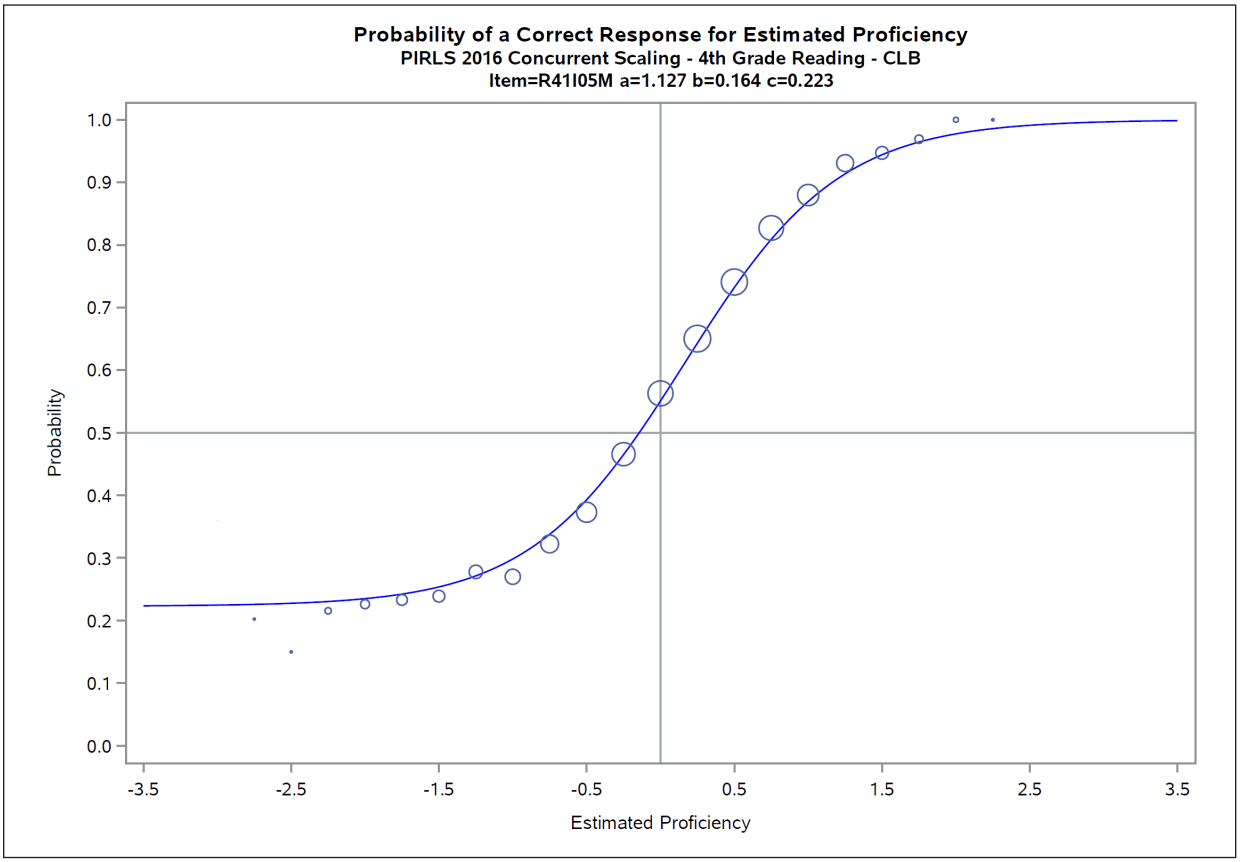
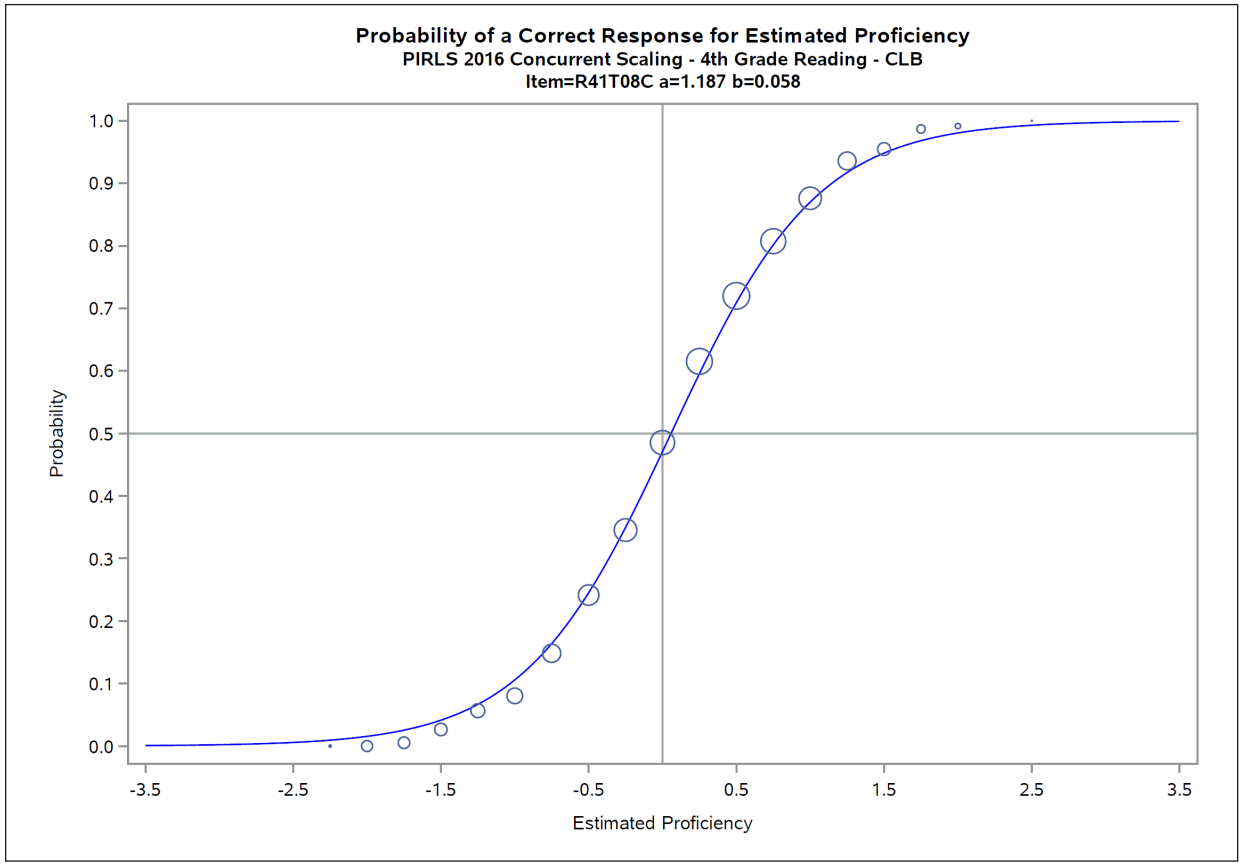
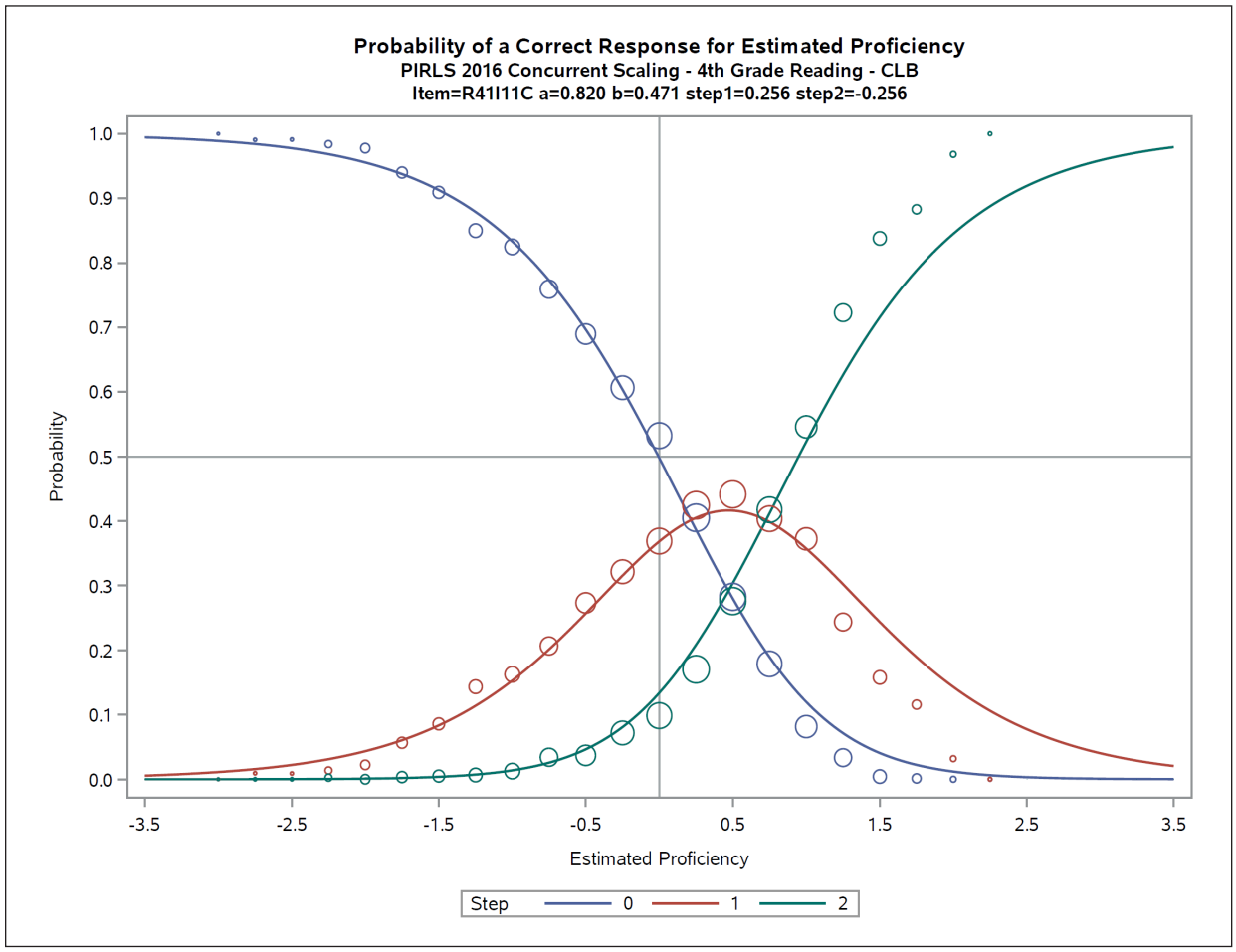


Exhibit 12.5: Example of Item Response Function for a Dichotomous Constructed Response Item from the PIRLS 2016 Assessment



The plot in Exhibit 12.6 shows the empirical and fitted item response functions for a polytomous item (scored 0, 1, or 2). As for the dichotomous item plots, the horizontal axis represents the proficiency scale, but in this example the vertical axis represents the probability of having a response in a given response category. The fitted curves based on the estimated item parameters are shown as solid lines and again the empirical results are represented by circles. The interpretation of the circles is the same as in Exhibits 12.4 and 12.5. The curve starting at the top left of the chart plots the probability of a score of zero on the item. This probability decreases as proficiency increases. The bell-shaped curve shows the probability of a score of one point—partial credit, starting low for low-ability students, reaching a maximum for medium-ability students, and decreasing for high-ability students. The curve ending at the top right corner of the chart shows the probability of a score of two points—full credit, starting low for low-ability students and increasing as proficiency increases.

Exhibit 12.6: Example of Item Response Function for a Polytomous Constructed Response Item from the PIRLS 2016 Assessment



Variables for Conditioning the PIRLS Assessment Data

Conditioning is the practice of using all available students' context information to improve the reliability of the estimated student proficiency scores. Ideally, all context data would be included in the conditioning model, but because PIRLS has so many student context variables that could be used in conditioning, the TIMSS & PIRLS International Study Center follows the practice established by NAEP and followed by other large-scale studies of using principal components analysis to reduce the number of variables while explaining most of their common variance. Principal components for the PIRLS student context variables (including parent context variables) were constructed as follows:

- For categorical variables (questions with a small number of fixed response options), a dummy coded variable was created for each response option, with a value of one if

the option is chosen and zero otherwise. If a student omitted or was not administered a particular question, all dummy coded variables associated with that question were assigned the value zero.

- Background variables with numerous response options (such as year of birth) were recoded using criterion scaling.[4] This was done by replacing the response option with the mean interim achievement score of all students choosing that option. Criterion scaling maximizes the correlation between the scaled variable and achievement. For PIRLS, the interim achievement score was the reading EAP scores produced from the item calibrations.
- Separately for each country, all the dummy-coded and criterion-scaled variables were included in a principal components analysis. Those principal components accounting for up to 90 percent of the variance of the context variables were retained for use as conditioning variables.[5] Because the principal components analysis was performed separately for each country, different numbers of principal components were required to account for 90% of the common variance in each country's context variables.

In addition to the principal components, student gender (dummy coded), the language of the test (dummy coded), an indicator of the classroom in the school to which a student belongs (criterion scaled), and an optional country-specific variable (dummy coded) were included as primary conditioning variables, thereby accounting for most of the variance between students and preserving the between-classroom and within-classroom variance structure in the scaling model. Exhibit 12.7 provides details on the conditioning models used for proficiency estimation in PIRLS 2016.

4 The process of generating criterion-scaled variables is described in Beaton (1969).

5 The number of principal components retained is limited to no more than 5% of a country's student sample size, thereby possibly reducing the percentage of variance accounted for, to avoid over-specification of the conditioning model.

Exhibit 12.7: Conditioning Models for PIRLS 2016 Achievement Scales

Country	2016				2011			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Australia	2	539	278	90	2	545	286	90
Austria	2	544	218	79	2	543	233	80
Azerbaijan	3	533	299	89	2	546	244	80
Bahrain	3	545	274	85	—	—	—	—
Belgium (Flemish)	2	541	259	85	—	—	—	—
Belgium (French)	2	526	231	82	2	548	186	73
Bulgaria	2	529	214	81	2	527	263	88
Canada	5	521	293	90	6	540	305	90
Chile	2	518	214	79	—	—	—	—
Chinese Taipei	2	542	216	79	2	553	214	77
Czech Republic	2	536	276	88	2	551	227	80
Denmark	2	545	175	71	2	555	229	78
England	2	248	145	90	2	243	137	90
Finland	3	544	244	84	3	550	232	81
France	2	543	238	82	2	547	221	78
Georgia	3	545	287	88	2	551	239	80
Germany	2	541	197	77	2	552	200	75
Hong Kong SAR	2	541	167	71	2	555	193	73
Hungary	2	521	231	82	2	539	260	85
Iran, Islamic Rep. of	2	545	219	80	2	555	287	86
Ireland	2	545	230	82	2	549	226	80
Israel	3	507	202	78	3	525	209	78
Italy	2	539	197	74	3	551	209	75
Kazakhstan	3	527	246	82	—	—	—	—
Latvia	3	545	207	77	—	—	—	—
Lithuania	4	524	215	79	2	547	233	80
Macao SAR	4	545	202	75	—	—	—	—
Malta	2	537	182	71	2	555	177	69
Morocco	2	545	274	86	2	549	324	90
Netherlands	2	539	210	81	2	546	199	76
New Zealand	7	515	278	90	8	549	282	88
Northern Ireland	2	507	184	79	2	544	179	75
Norway (5)	3	526	211	78	—	—	—	—

Exhibit 12.7: Conditioning Models for PIRLS 2016 Achievement Scales (Continued)

Country	2016				2011			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Oman	3	545	317	90	3	553	323	90
Poland	2	532	220	81	—	—	—	—
Portugal	2	544	232	81	2	542	204	77
Qatar	3	542	307	90	3	544	206	75
Russian Federation	2	521	228	81	2	527	223	80
Saudi Arabia	3	545	237	80	3	544	225	78
Singapore	2	545	300	90	2	555	307	90
Slovak Republic	3	545	272	87	3	547	281	88
Slovenia	2	540	224	81	2	547	225	80
Spain	8	538	303	90	7	544	304	90
Sweden	2	521	226	82	2	547	231	81
Trinidad and Tobago	2	513	208	77	2	525	197	75
United Arab Emirates	6	545	316	90	5	541	317	90
United States	9	250	147	90	9	237	141	90
Benchmarking Participants								
Buenos Aires, Argentina	2	530	219	80	—	—	—	—
Ontario, Canada	3	519	213	81	3	540	228	82
Quebec, Canada	3	519	158	70	3	540	212	78
Norway (4)	3	527	217	79	3	553	159	69
Moscow City, Russian Fed.	2	521	214	78	—	—	—	—
Eng/Afr/Zulu - RSA (5)	4	539	264	85	—	—	—	—
Andalusia, Spain	2	538	208	77	2	541	216	77
Madrid, Spain	2	537	189	73	—	—	—	—
Abu Dhabi, UAE	3	545	209	76	3	541	207	75
Dubai, UAE	4	545	306	90	3	541	303	90

Generating IRT Proficiency Scores for the PIRLS Assessment Data

Educational Testing Service's DGROUP program (Rogers, Tang, Lin, & Kandathil, 2006) was used to generate the IRT proficiency scores. This program takes as input the students' responses to the items they were given, the item parameters estimated at the calibration stage, and the conditioning variables, and generates as output the plausible values that represent student proficiency.

A useful feature of DGROUP is its ability to perform multi-dimensional scaling using the responses to all items across the proficiency scales and the correlations among the scales to improve the reliability of each individual scale. The multi-dimensional scaling feature of DGROUP also was used to generate proficiency scores for the PIRLS 2016 domains. The estimation of proficiency scores for the purposes for reading and the processes of comprehension relied on multidimensional IRT models using the item parameters estimated for the overall reading scale as well the same conditioning variables. PIRLS 2016 used two two-dimensional scaling models, one to estimate proficiency scores for the two purposes for reading and a second for the two processes of comprehension.

In addition to generating plausible values for the overall reading scale from the 2016 assessment data, the item parameters estimated at the calibration stage also were used to generate plausible values for the PIRLS 2011 assessment for the countries included in the concurrent calibration. These additional plausible values were used to establish the linear transformation necessary to place the 2016 assessment data on the PIRLS reading trend scale.

Transforming the Overall Scores to Measure Trends

To provide results for the PIRLS 2016 assessment on the PIRLS achievement scales, the 2016 proficiency scores (plausible values) for overall reading had to be transformed to the PIRLS reporting metric. This was accomplished through a set of linear transformations as part of the concurrent calibration approach. These linear transformations were given by:

$$PV^{*}_{k,i} = A_{k,i} + B_{k,i} \times PV_{k,i} \qquad (12.1)$$

where

$PV_{k,i}$ is the PIRLS 2016 plausible value i of scale k prior to transformation;
$PV^{*}_{k,i}$ is the PIRLS 2016 plausible value i of scale k after transformation; and
$A_{k,i}$ and $B_{k,i}$ are the linear transformation constants.

The linear transformation constants were obtained by first computing the international means and standard deviations of the proficiency scores for the overall reading scale using the plausible values produced in 2011 based on the 2011 item calibrations for the trend countries. These were the plausible values published in 2011. Next, the same calculations were done using the plausible values

from the re-scaled PIRLS 2011 assessment data based on the 2016 concurrent item calibration for the same set of countries. From these calculations, the linear transformation constants were defined as:

$$B_{k,i} = \sigma_{k,i} / \sigma^*_{k,i} \tag{12.2}$$

$$A_{k,i} = \mu_{k,i} - B_{k,i} \cdot \mu^*_{k,i} \tag{12.3}$$

where

$\mu_{k,i}$ is the international mean of scale k based on plausible value i published in 2011;

$\mu^*_{k,i}$ is the international mean of scale k based on plausible value i from the 2011 assessment based on the 2016 concurrent calibration;

$\sigma_{k,i}$ is the international standard deviation of scale k based on plausible value i published in 2011;

$\sigma^*_{k,i}$ is the international standard deviation of scale k based on plausible value i from the 2011 assessment based on the 2016 concurrent calibration.

There are five sets of transformation constants for the PIRLS reading scale, one for each plausible value. The trend countries contributed equally in the calculation of these transformation constants. Exhibit 12.8 shows the PIRLS 2016 transformation constants for overall reading.

Exhibit 12.8: Linear Transformation Constants for PIRLS 2016 Achievement Scales

Overall Reading	PIRLS 2011 Published Scores		PIRLS 2011 Re-scaled Scores		$A_{k,i}$	$B_{k,i}$
	Mean	Standard Deviation	Mean	Standard Deviation		
PV1	514.88796	93.40789	-0.02153	0.96698	516.96808	96.59763
PV2	514.33588	94.16192	-0.01873	0.96533	516.16294	97.54392
PV3	514.10484	93.95296	-0.01702	0.96329	515.76531	97.53376
PV4	514.09822	94.15851	-0.01852	0.96502	515.90514	97.57133
PV5	514.19052	93.93593	-0.01874	0.96576	516.01365	97.26663

These linear transformation constants were applied to the overall reading proficiency scores and for all participating countries and benchmarking participants. This provided student achievement scores for the PIRLS 2016 assessment that are directly comparable to the scores from all previous assessments.

The linear transformation constants for overall reading also were applied to the scales for the purposes for reading and the processes of comprehension. In this approach to measuring trends in the purposes and processes, achievement changes over time are established in the context of achievement in overall reading. Trends are not established separately for each purpose or process; rather differential changes in performance in the domains are considered in the context of trends in overall reading.

Scaling the PIRLS Literacy 2016 Achievement Data

Launched in 2011 as prePIRLS, PIRLS Literacy 2016 is a reading assessment intended for populations of readers that would find the PIRLS reading assessment too challenging. Although a less demanding assessment, PIRLS Literacy was designed to allow the reading achievement of participating countries to be reported on the PIRLS reading trend scale. To that end, PIRLS and PIRLS Literacy in 2016 shared four passages to establish a psychometric link between the two assessments. Two shared passages were PIRLS passages with their usual structure of a text accompanied by a set of items related to that text. Two shared passages were PIRLS Literacy passages with their items interspersed within the accompanying text.

Exhibit 12.9 shows the number of items present in the PIRLS Literacy 2016 assessment by item type and domain. There was a total of 183 items in the PIRLS Literacy assessment, 59 of them shared with the PIRLS reading assessment.

Exhibit 12.9: PIRLS Literacy 2016 Items for Calibration

Item Type	Points	PIRLS Literacy Shared Items		PIRLS Literacy Unique Items		Total	
		Items	Points	Items	Points	Items	Points
Multiple-Choice	1	29	29	61	61	90	90
Constructed Response	1	18	18	49	49	67	67
	2	11	22	12	24	23	46
	3	1	3	2	6	3	9
Total		**59**	**72**	**124**	**140**	**183**	**212**

PIRLS Literacy 2016 Items for Calibration by Reading Purposes and Comprehension Processes

Purposes for Reading	PIRLS Literacy Shared Items		PIRLS Literacy Unique Items		Total	
	Items	Points	Items	Points	Items	Points
Literary Experience	30	36	63	71	93	107
Acquire and Use Information	29	36	61	69	90	105
Processes of Comprehension	**PIRLS Literacy Shared Items**		**PIRLS Literacy Unique Items**		**Total**	
	Items	Points	Items	Points	Items	Points
Retrieving and Straightforward Inferencing	40	44	94	101	134	145
Interpreting, Integrating, and Evaluating	19	28	30	39	49	67
Total	**59**	**72**	**124**	**140**	**183**	**212**

Much like the normal PIRLS scaling procedure, the PIRLS Literacy scaling approach involved the same four tasks of calibrating the achievement items, creating principal components for conditioning, generating proficiency scores, and placing these proficiency scores on the PIRLS reading reporting scale. Exhibit 12.10 shows the sample sizes for scaling the PIRLS Literacy data. A total of six countries participated and all were included in the item calibration—including data from Denmark's benchmarking participation in PIRLS Literacy 2016 at the 3rd grade.

Exhibit 12.10: Sample Sizes for PIRLS Literacy 2016 Achievement Scales

Country	Item Calibration	Proficiency Estimation
Egypt	6,957	6,957
Iran, Islamic Rep. of	4,381	4,381
Kuwait	4,609	4,609
Morocco	5,453	5,453
South Africa	12,810	12,810
Benchmarking Participants		
Denmark (3)	3,600	3,600
Total	**37,810**	**37,810**

The item calibration step was based on a straightforward calibration of the PIRLS Literacy 2016 achievement items from the six participating countries. The item parameters for the PIRLS Literacy items were placed on the PIRLS reading metric by fixing the parameters of the items in the four shared passages to the values estimated from the PIRLS 2016 concurrent calibration. The item parameters estimated from the PIRLS Literacy 2016 item calibration are presented in Appendix 12B. The 59 link items, whose item parameters were fixed, are marked with asterisks.

The conditioning for PIRLS Literacy 2016 was done in exactly the same way as for PIRLS, as was the estimation of proficiency scores using the DGROUP software. This included overall reading scores for the PIRLS Literacy countries and scores for the PIRLS purposes for reading and processes of comprehension. Exhibit 12.11 provides details on the conditioning models used for the PIRLS Literacy 2016 proficiency estimation.

Exhibit 12.11: PIRLS Literacy 2016 Conditioning Models for Proficiency Estimation

Country	2016			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Egypt	2	545	304	90
Iran, Islamic Rep. of	2	545	219	80
Kuwait	3	535	230	80
Morocco	2	545	272	85
South Africa	12	539	323	90
Benchmarking Participants				
Denmark (3)	2	545	180	72

The final step in the process consisted of placing students' performance on the PIRLS Literacy 2016 assessment on the PIRLS reading reporting scale. This was done by applying the appropriate linear transformation to the estimated proficiency scores. The PIRLS Literacy 2016 item calibration resulted in item parameters on the same metric as the PIRLS 2016 concurrent calibration—by fixing the parameters of the 59 link items. Thus, placing the PIRLS Literacy 2016 achievement scores on the PIRLS reporting scale was accomplished by using the PIRLS 2016 reading linear transformation constants, as presented in Exhibit 12.8. These linear transformation constants were applied to the PIRLS Literacy 2016 overall reading achievement scores, as well as the achievement scores on the purposes for reading and the processes of comprehension.

In 2011, PIRLS Literacy's predecessor prePIRLS was reported as its own scale, although its item parameters were estimated on the same item parameter metric, capitalizing on Colombia's participation in both PIRLS and prePIRLS in 2011. However, with South Africa having participated in both prePIRLS in 2011 and PIRLS Literacy in 2016, there was a need to place their 2011 results on the PIRLS trend scale. To that end, it was necessary to re-transform their achievement scores—overall reading, as well as the purposes and processes—using the PIRLS 2011 linear transformation constants, as given in Exhibit 15 of the Scaling the TIMSS and PIRLS Achievement 2011 Data chapter of Methods and Procedures in TIMSS and PIRLS 2011.

Scaling the ePIRLS 2016 Achievement Data

ePIRLS 2016 is a new computer-based assessment of online informational reading, consisting of five tasks, designed to assess fourth grade students' ability to use the internet in a school context. With ePIRLS designed as an extension of PIRLS reading, students participating in ePIRLS 2016

were expected also to participate in PIRLS 2016. Thus, 14 countries and two benchmarking entities participated in both PIRLS and ePIRLS in 2016. Exhibit 12.12 lists the ePIRLS participants and their effective sample sizes across both PIRLS and ePIRLS assessments in 2016.

Exhibit 12.12: PIRLS 2016 and ePIRLS 2016 Sample Sizes

Country	PIRLS Sample Size	ePIRLS Sample Size	Percentage Overlap	ePIRLS Students not in PIRLS	Percentage not in PIRLS
Canada	18,245	8,871	48.6	261	2.9
Chinese Taipei	4,326	4,299	99.4	32	0.7
Denmark	3,508	2,506	71.4	120	4.6
Georgia	5,741	5,557	96.8	100	1.8
Ireland	4,607	2,473	53.7	82	3.2
Israel	4,041	3,798	94.0	135	3.4
Italy	3,940	3,767	95.6	95	2.5
Norway (5)	4,232	3,610	85.3	104	2.8
Portugal	4,642	4,558	98.2	78	1.7
Singapore	6,488	6,320	97.4	100	1.6
Slovenia	4,499	4,303	95.6	67	1.5
Sweden	4,525	3,879	85.7	109	2.7
United Arab Emirates	16,471	15,566	94.5	441	2.8
United States	4,425	4,090	92.4	16	0.4
Total	**89,690**	**73,597**	**82.1**	**1,740**	**1.9**
Benchmarking Participants					
Abu Dhabi, UAE	4,188	3,980	95.0	86	2.1
Dubai, UAE	7,859	7,471	95.1	155	2.0

In general, ePIRLS 2016 participants were successful in having nearly all their sampled students participate in both assessments, with a few notable exceptions. In Canada, some provinces did not take part in ePIRLS and a subsample of Quebec's PIRLS schools participated in ePIRLS. In Ireland, because of limitations in the number of computers available in many schools, random subsamples of PIRLS students participated in ePIRLS. In Denmark, Norway, and Sweden, some PIRLS schools were unable to participate in ePIRLS, generally arising from the absence of compatible computers for the ePIRLS assessment. It is worth pointing out that a small proportion of students—less than 2% internationally—took part in the ePIRLS assessment, but not in the PIRLS assessment. These students were removed from the ePIRLS samples. Thus, only students that participated in both PIRLS and ePIRLS assessments were retained in the ePIRLS samples.

Exhibit 12.13 shows the number of items present in the ePIRLS 2016 assessment by item type and process of comprehension. The exhibit also includes the PIRLS 2016 items since they were included in the ePIRLS item calibration. There was a total of 91 items in the ePIRLS 2016 assessment. The 175 PIRLS 2016 items were also included in the item calibration, with fixed item parameters from the PIRLS 2016 concurrent calibration.

Exhibit 12.13: ePIRLS 2016 Items for Calibration

Item Type	Points	ePIRLS 2016 Items		PIRLS 2016 Items		Total	
		Items	Points	Items	Points	Items	Points
Multiple-Choice	1	36	36	86	86	122	122
Constructed Response	1	37	37	49	49	86	86
	2	15	30	32	64	47	94
	3	3	9	8	24	11	33
Total		**91**	**112**	**175**	**223**	**266**	**335**

ePIRLS 2016 Items by Comprehension Process

Processes of Comprehension	ePIRLS 2016 Items		PIRLS 2016 Items		Total	
	Items	Points	Items	Points	Items	Points
Retrieving and Straightforward Inferencing	49	54	103	116	152	170
Interpreting, Integrating, and Evaluating	42	58	72	107	114	165
Total	**91**	**112**	**175**	**223**	**266**	**335**

The ePIRLS scaling methodology adopted the same four steps of calibration, conditioning, generating proficiency scores, and placing those scores on the PIRLS reading scale. All 14 ePIRLS countries were included in the item calibration, including their responses to the PIRLS and ePIRLS items. The item parameters for the ePIRLS 2016 items were placed on the PIRLS reading metric by fixing the parameters of the PIRLS 2016 items to the values estimated from the PIRLS 2016 concurrent calibration. The item parameters estimated from the ePIRLS 2016 item calibration are presented in Appendix 12C. Although the PIRLS 2016 items were included in the ePIRLS item calibration, they are not included in Appendix 12C as they are in every way identical to the parameters estimated for PIRLS 2016 and presented in Appendix 12A.

Exhibit 12.14 provides details on the conditioning models used for the ePIRLS 2016 proficiency estimation. Although ePIRLS used the same set of conditioning variables from the PIRLS student and parents questionnaires, the resulting conditioning matrices were not necessarily

identical to PIRLS since the ePIRLS samples sizes were not the same as the PIRLS sample sizes. The DGROUP software was used to estimate ePIRLS proficiency scores, including overall ePIRLS online informational reading scores and scores for the two PIRLS processes of comprehension.

Exhibit 12.14: ePIRLS 2016 Conditioning Models for Proficiency Estimation

Country	2016			
	Number of Primary Conditioning Variables	Number of Principal Components Available	Number of Principal Components Retained	Percentage of Variance Explained
Canada	5	521	279	90
Chinese Taipei	2	542	214	79
Denmark	2	545	125	62
Georgia	3	545	277	87
Ireland	2	545	123	62
Israel	3	507	189	76
Italy	2	539	188	73
Norway (5)	3	526	180	73
Portugal	2	544	227	80
Singapore	2	545	299	90
Slovenia	2	540	215	80
Sweden	2	521	193	77
United Arab Emirates	6	545	315	90
United States	9	250	147	90
Benchmarking Participants				
Abu Dhabi, UAE	3	545	199	75
Dubai, UAE	4	545	306	90

The final step in the process consisted of placing students' performance on the ePIRLS 2016 assessment on the PIRLS reading reporting scale. This was done by applying the appropriate linear transformation to the estimated proficiency scores. The ePIRLS 2016 item calibration resulted in item parameters on the same metric as the PIRLS reading metric—by fixing the parameters of all PIRLS 2016 items. Thus, placing the ePIRLS achievement scores on the PIRLS reporting scale was accomplished by using the PIRLS 2016 reading linear transformation constants, as presented in Exhibit 12.8. These linear transformation constants were applied to the ePIRLS 2016 overall online informational reading achievement scores, as well as the achievement scores on the two processes of comprehension.

References

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

Mazzeo, J., and von Davier, M. (2014) Linking scales in International Large-Scale Assessments. In L. Rutkowski, M. von Davier, & D. Rutkowski, *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis* (pp. 229-258). Boca Raton: Chapman & Hall/CRC.

Muraki, E., & Bock, R.D. (1991). *PARSCALE: Parameter scaling of rating data* [computer software]. Chicago, IL: Scientific Software, Inc.

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

Appendix 12A: PIRLS 2016 Item Parameters from Concurrent Calibration

Item Parameters from PIRLS 2016 Concurrent Item Calibration

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
Items Released in 2011:						
R21E01M	1.375 (0.091)	-1.108 (0.077)	0.378 (0.035)			
R21E02M	1.143 (0.071)	-0.344 (0.061)	0.287 (0.027)			
R21E03M	0.552 (0.045)	-0.337 (0.127)	0.153 (0.039)			
R21E04M	1.452 (0.091)	-1.076 (0.067)	0.330 (0.033)			
R21E05C	0.619 (0.017)	-0.495 (0.025)		-0.539 (0.051)	0.539 (0.044)	
R21E06M	1.323 (0.076)	-0.228 (0.046)	0.242 (0.023)			
R21E07C	0.657 (0.021)	-0.243 (0.023)		0.180 (0.042)	-0.180 (0.036)	
R21E08M	1.321 (0.081)	0.477 (0.032)	0.162 (0.015)			
R21E09C	0.534 (0.021)	0.735 (0.031)		0.534 (0.042)	-0.534 (0.053)	
R21E10C	0.964 (0.035)	-0.174 (0.025)				
R21E11M	0.901 (0.071)	0.266 (0.065)	0.249 (0.026)			
R21E12C	0.780 (0.026)	0.173 (0.019)		0.315 (0.032)	-0.315 (0.032)	
R31P01M	1.106 (0.057)	-0.484 (0.051)	0.149 (0.025)			
R31P02C	0.856 (0.032)	-0.776 (0.035)				
R31P03C	1.095 (0.038)	-0.529 (0.025)				
R31P04M	0.990 (0.067)	0.583 (0.040)	0.130 (0.017)			
R31P05C	0.649 (0.019)	0.421 (0.020)		-0.310 (0.040)	0.310 (0.042)	
R31P06C	1.353 (0.046)	-0.613 (0.023)				
R31P07C	0.941 (0.024)	-0.117 (0.016)		-0.137 (0.031)	0.137 (0.028)	
R31P08M	1.090 (0.068)	-0.335 (0.063)	0.263 (0.028)			
R31P09C	1.199 (0.041)	-0.485 (0.024)				
R31P10M	1.769 (0.095)	-0.520 (0.039)	0.240 (0.023)			
R31P11M	1.152 (0.063)	-0.226 (0.048)	0.171 (0.023)			
R31P12M	1.342 (0.076)	0.133 (0.036)	0.182 (0.019)			
R31P13M	1.325 (0.072)	-0.753 (0.056)	0.221 (0.030)			
R31P14C	1.184 (0.041)	0.028 (0.020)				
R31P15C	0.630 (0.023)	0.397 (0.022)		0.173 (0.038)	-0.173 (0.041)	
R31P16C	0.783 (0.036)	0.744 (0.033)				
R21N01M	0.852 (0.059)	-0.640 (0.099)	0.281 (0.037)			
R21N02C	0.780 (0.030)	-0.494 (0.033)				

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R21N03C	0.747 (0.028)	1.061 (0.029)		0.353 (0.031)	-0.353 (0.047)	
R21N04M	1.175 (0.074)	0.207 (0.042)	0.207 (0.020)			
R21N05M	1.610 (0.093)	-0.856 (0.051)	0.276 (0.029)			
R21N06M	1.457 (0.076)	-0.475 (0.042)	0.201 (0.023)			
R21N07M	1.074 (0.066)	-0.043 (0.052)	0.205 (0.024)			
R21N08C	0.933 (0.034)	-0.269 (0.026)				
R21N09M	1.178 (0.072)	-0.337 (0.058)	0.270 (0.027)			
R21N10M	0.878 (0.073)	0.249 (0.073)	0.284 (0.028)			
R21N11C	0.555 (0.016)	-0.010 (0.023)		-0.555 (0.050)	0.555 (0.048)	
R21N12C	0.636 (0.022)	0.115 (0.021)		0.080 (0.041)	-0.080 (0.039)	
R31G01M	1.116 (0.066)	-0.513 (0.064)	0.257 (0.030)			
R31G02C	0.680 (0.028)	-0.160 (0.031)				
R31G03M	1.100 (0.067)	-0.303 (0.060)	0.253 (0.027)			
R31G04C	0.863 (0.038)	0.982 (0.036)				
R31G05M	1.178 (0.091)	0.481 (0.045)	0.288 (0.020)			
R31G06M	1.019 (0.059)	-0.309 (0.059)	0.195 (0.027)			
R31G07M	1.101 (0.066)	0.088 (0.045)	0.180 (0.021)			
R31G08CZ	0.792 (0.028)	0.977 (0.026)		0.264 (0.029)	-0.264 (0.042)	
R31G09M	0.877 (0.061)	0.079 (0.064)	0.197 (0.026)			
R31G10C	0.993 (0.038)	0.566 (0.024)				
R31G11M	1.612 (0.107)	0.336 (0.034)	0.302 (0.017)			
R31G12C	0.465 (0.018)	1.639 (0.059)		-0.863 (0.063)	0.863 (0.086)	
R31G13CZ	0.819 (0.019)	0.157 (0.013)		-0.280 (0.039)	0.108 (0.044)	0.171 (0.036)
R31G14M	1.312 (0.088)	0.359 (0.039)	0.241 (0.019)			
Items Common in 2011 and 2016:						
R11F01M	1.334 (0.049)	-0.627 (0.034)	0.148 (0.018)			
R11F02M	0.666 (0.038)	-0.848 (0.111)	0.243 (0.037)			
R11F03M	0.920 (0.039)	-0.666 (0.054)	0.157 (0.024)			
R11F04M	1.307 (0.053)	-0.831 (0.044)	0.228 (0.023)			
R11F05M	0.940 (0.045)	-0.255 (0.052)	0.217 (0.022)			
R11F06C	0.776 (0.023)	-0.152 (0.021)				
R11F07C	0.503 (0.010)	0.375 (0.018)		-0.896 (0.041)	0.896 (0.043)	
R11F08C	1.149 (0.029)	-0.328 (0.017)				
R11F09C	1.011 (0.022)	-0.627 (0.015)		0.074 (0.027)	-0.074 (0.020)	

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R11F10C	0.846 (0.026)	-1.419 (0.039)				
R11F11M	0.739 (0.045)	0.208 (0.061)	0.192 (0.023)			
R11F12C	0.618 (0.014)	0.642 (0.017)		-0.471 (0.032)	0.471 (0.036)	
R11F13M	1.124 (0.054)	-0.199 (0.046)	0.270 (0.021)			
R21Y01M	1.097 (0.055)	0.154 (0.038)	0.253 (0.017)			
R21Y02M	1.649 (0.070)	-0.204 (0.029)	0.288 (0.016)			
R21Y03C	0.815 (0.025)	0.564 (0.021)				
R21Y04M	1.273 (0.056)	0.093 (0.031)	0.222 (0.015)			
R21Y05M	1.721 (0.070)	0.086 (0.022)	0.226 (0.013)			
R21Y06M	1.533 (0.062)	0.042 (0.025)	0.209 (0.014)			
R21Y07M	0.792 (0.037)	-1.011 (0.079)	0.182 (0.030)			
R21Y08M	1.360 (0.058)	-0.271 (0.035)	0.261 (0.018)			
R21Y09C	0.956 (0.020)	-0.551 (0.015)		0.078 (0.027)	-0.078 (0.020)	
R21Y10C	0.749 (0.024)	0.574 (0.023)				
R21Y11M	1.411 (0.065)	0.035 (0.032)	0.284 (0.016)			
R21Y12C	0.706 (0.012)	-0.001 (0.014)		-1.154 (0.039)	1.154 (0.038)	
R21Y13C	0.760 (0.017)	0.378 (0.011)		0.594 (0.026)	-0.219 (0.028)	-0.375 (0.031)
R21Y14C	0.576 (0.013)	0.222 (0.016)		-0.549 (0.036)	0.549 (0.036)	
R31M01M	1.451 (0.062)	-0.877 (0.043)	0.268 (0.023)			
R31M02C	1.218 (0.033)	-0.957 (0.023)				
R31M03M	1.330 (0.057)	-0.004 (0.031)	0.228 (0.016)			
R31M04C	0.560 (0.020)	0.357 (0.028)				
R31M05M	1.551 (0.083)	0.112 (0.034)	0.424 (0.015)			
R31M06M	1.159 (0.063)	0.368 (0.036)	0.283 (0.016)			
R31M07M	1.619 (0.068)	-0.350 (0.031)	0.276 (0.017)			
R31M08M	1.382 (0.058)	-0.498 (0.038)	0.265 (0.020)			
R31M09C	0.759 (0.015)	-0.030 (0.017)		1.143 (0.027)	-1.143 (0.024)	
R31M10C	0.623 (0.021)	0.413 (0.025)				
R31M11M	0.854 (0.043)	-0.692 (0.075)	0.262 (0.029)			
R31M12M	1.196 (0.050)	0.106 (0.030)	0.162 (0.015)			
R31M13M	2.100 (0.089)	-0.642 (0.027)	0.256 (0.018)			
R31M14M	2.283 (0.087)	-0.195 (0.019)	0.197 (0.013)			
R31M15M	1.298 (0.057)	0.065 (0.031)	0.218 (0.016)			
R31M16C	1.207 (0.031)	0.058 (0.015)				
R31M17CZ	0.612 (0.014)	0.003 (0.013)		0.071 (0.038)	0.228 (0.037)	-0.299 (0.032)

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R11L01M	0.532 (0.027)	-2.275 (0.169)	0.146 (0.053)			
R11L02M	0.749 (0.056)	0.685 (0.054)	0.241 (0.020)			
R11L03C	0.616 (0.020)	-0.474 (0.029)				
R11L04C	0.667 (0.012)	0.418 (0.015)		1.643 (0.030)	-1.051 (0.032)	-0.592 (0.043)
R11L05M	1.186 (0.057)	0.352 (0.030)	0.206 (0.014)			
R11L06C	0.656 (0.021)	0.191 (0.023)				
R11L07M	0.772 (0.045)	0.474 (0.045)	0.154 (0.018)			
R11L08C	0.801 (0.019)	0.612 (0.015)		0.703 (0.021)	-0.703 (0.027)	
R11L09M	0.963 (0.043)	-0.809 (0.061)	0.226 (0.026)			
R11L10C	0.732 (0.019)	0.681 (0.016)		0.231 (0.024)	-0.231 (0.029)	
R11L11M	0.912 (0.042)	-0.354 (0.052)	0.189 (0.022)			
R11L12C	0.735 (0.017)	0.509 (0.016)		0.810 (0.023)	-0.810 (0.028)	
R21K01C	0.422 (0.013)	-0.891 (0.033)		0.186 (0.052)	-0.186 (0.039)	
R21K02C	0.807 (0.023)	-0.559 (0.025)				
R21K03M	1.004 (0.047)	0.081 (0.039)	0.184 (0.018)			
R21K04M	1.062 (0.096)	0.979 (0.045)	0.391 (0.014)			
R21K05C	0.969 (0.026)	0.137 (0.017)				
R21K06M	1.489 (0.067)	0.052 (0.029)	0.281 (0.015)			
R21K07C	0.682 (0.016)	0.143 (0.015)		0.119 (0.027)	-0.119 (0.027)	
R21K08M	0.994 (0.052)	0.354 (0.037)	0.197 (0.017)			
R21K09M	1.203 (0.056)	-0.010 (0.037)	0.246 (0.018)			
R21K10C	0.785 (0.017)	0.778 (0.015)		-0.397 (0.027)	0.397 (0.031)	
R21K11M	1.070 (0.056)	0.260 (0.039)	0.240 (0.017)			
R21K12C	0.576 (0.013)	-0.110 (0.014)		0.446 (0.040)	-0.084 (0.037)	-0.362 (0.034)
R31W01C	0.718 (0.017)	-0.584 (0.018)		0.243 (0.032)	-0.243 (0.024)	
R31W02C	0.800 (0.017)	0.278 (0.013)		-0.107 (0.024)	0.107 (0.025)	
R31W03M	1.347 (0.052)	-0.063 (0.027)	0.162 (0.014)			
R31W04C	0.842 (0.024)	-0.687 (0.026)				
R31W05M	1.264 (0.068)	0.497 (0.030)	0.257 (0.014)			
R31W06M	0.753 (0.034)	-0.999 (0.080)	0.147 (0.031)			
R31W07CZ	0.879 (0.017)	0.509 (0.010)		-0.079 (0.025)	0.169 (0.029)	-0.090 (0.027)
R31W08M	1.355 (0.063)	-0.093 (0.037)	0.307 (0.018)			
R31W09M	0.951 (0.054)	0.565 (0.036)	0.178 (0.016)			
R31W10M	1.289 (0.056)	0.320 (0.026)	0.164 (0.013)			
R31W11C	1.467 (0.038)	0.551 (0.013)				

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R31W12M	1.514 (0.081)	0.706 (0.023)	0.219 (0.011)			
R31W13C	0.862 (0.028)	0.791 (0.023)				
Items Introduced in 2016:						
L21B01C	0.677 (0.036)	-2.530 (0.108)				
L21B02M	0.958 (0.067)	-2.353 (0.150)	0.101 (0.072)			
L21B03M *	0.721 (0.048)	-2.406 (0.130)	0.250 (0.000)			
L21B04C	0.706 (0.032)	-1.260 (0.057)				
L21B05M	0.883 (0.061)	-1.315 (0.127)	0.223 (0.054)			
L21B06M *	0.440 (0.033)	-1.847 (0.144)	0.250 (0.000)			
L21B07C	0.745 (0.033)	-1.160 (0.052)				
L21B08C	0.840 (0.036)	-1.359 (0.053)				
L21B09C	0.855 (0.036)	-1.140 (0.047)				
L21B10M	0.660 (0.057)	-0.338 (0.133)	0.192 (0.046)			
L21B11M	0.979 (0.062)	-1.224 (0.103)	0.186 (0.048)			
L21B12M	0.738 (0.055)	-2.202 (0.212)	0.127 (0.091)			
L21B13C	0.542 (0.022)	-0.967 (0.040)		0.693 (0.067)	-0.693 (0.044)	
L21B14M	1.075 (0.063)	-0.597 (0.066)	0.149 (0.032)			
L21B15C	0.419 (0.026)	0.318 (0.053)				
L21B16C	0.435 (0.019)	-0.185 (0.035)		0.092 (0.065)	-0.092 (0.059)	
L21B17C	0.705 (0.027)	-0.414 (0.027)		0.684 (0.048)	-0.684 (0.035)	
R41H01M	0.947 (0.101)	-0.667 (0.166)	0.582 (0.044)			
R41H02M	1.058 (0.066)	-0.743 (0.079)	0.196 (0.038)			
R41H03C	1.172 (0.044)	0.257 (0.022)				
R41H04C	0.671 (0.043)	1.644 (0.083)				
R41H05M	1.030 (0.082)	0.131 (0.070)	0.286 (0.029)			
R41H06C	0.700 (0.022)	-0.264 (0.024)		-0.089 (0.045)	0.089 (0.039)	
R41H07M	0.895 (0.085)	0.873 (0.053)	0.151 (0.022)			
R41H08C	0.691 (0.042)	1.492 (0.071)				
R41H09M	0.649 (0.064)	0.272 (0.110)	0.166 (0.040)			
R41H10M	1.307 (0.086)	-0.179 (0.056)	0.263 (0.028)			
R41H11M	1.402 (0.094)	-0.537 (0.065)	0.311 (0.034)			
R41H12M	1.350 (0.104)	0.002 (0.061)	0.369 (0.027)			
R41H13C	0.541 (0.018)	0.753 (0.023)		-0.089 (0.054)	0.270 (0.063)	-0.181 (0.068)
R41H14C	0.990 (0.042)	0.307 (0.025)				

* Items with fixed guessing parameters.

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R41H15C	1.250 (0.050)	-0.385 (0.028)				
R41H16C	1.027 (0.046)	0.577 (0.026)				
R41O01M	0.942 (0.065)	-0.993 (0.109)	0.254 (0.046)			
R41O02C	0.859 (0.036)	-0.965 (0.044)				
R41O03C	1.020 (0.046)	0.868 (0.030)				
R41O04C	0.595 (0.024)	1.145 (0.038)		-0.160 (0.046)	0.160 (0.061)	
R41O05C	0.630 (0.020)	0.228 (0.022)		-0.260 (0.046)	0.260 (0.045)	
R41O06M	1.437 (0.097)	-0.212 (0.055)	0.315 (0.028)			
R41O07C	0.629 (0.022)	-0.953 (0.037)		0.036 (0.062)	-0.036 (0.043)	
R41O08C	0.847 (0.036)	-0.064 (0.029)				
R41O09C	1.205 (0.045)	-0.059 (0.023)				
R41O10C	0.815 (0.026)	0.236 (0.019)		-0.006 (0.036)	0.006 (0.035)	
R41O11M	1.365 (0.096)	0.096 (0.050)	0.295 (0.025)			
R41O12M	1.283 (0.074)	-0.341 (0.051)	0.169 (0.027)			
R41O13C	0.567 (0.014)	0.348 (0.018)		-1.197 (0.072)	0.794 (0.079)	0.403 (0.057)
L21E01C	0.796 (0.047)	-3.130 (0.125)				
L21E02C	0.771 (0.038)	-2.116 (0.083)				
L21E03M	1.343 (0.084)	-0.521 (0.059)	0.270 (0.029)			
L21E04M	1.106 (0.074)	-0.378 (0.070)	0.262 (0.031)			
L21E05M	1.118 (0.076)	-1.568 (0.110)	0.242 (0.052)			
L21E06M	0.936 (0.081)	-2.465 (0.221)	0.304 (0.097)			
L21E07C	0.463 (0.018)	-0.810 (0.042)		-0.018 (0.071)	0.018 (0.055)	
L21E08M	1.023 (0.064)	-0.749 (0.081)	0.201 (0.036)			
L21E09M	0.620 (0.085)	0.464 (0.145)	0.340 (0.041)			
L21E10C	0.859 (0.047)	-2.716 (0.101)				
L21E11M	1.205 (0.074)	-0.817 (0.071)	0.226 (0.035)			
L21E12C	1.116 (0.057)	-2.136 (0.066)				
L21E13C	0.528 (0.027)	-0.275 (0.046)				
L21E14C	0.493 (0.020)	0.186 (0.032)		0.852 (0.053)	-0.852 (0.053)	
L21E15C	0.795 (0.040)	-2.219 (0.087)				
L21E16C	0.706 (0.032)	-0.959 (0.051)				
L21E17M	1.047 (0.065)	-0.467 (0.069)	0.171 (0.032)			
R41I01C	0.793 (0.036)	-1.522 (0.063)				
R41I02M	1.034 (0.085)	0.623 (0.048)	0.185 (0.021)			
R41I03C	0.560 (0.022)	0.201 (0.026)		0.330 (0.047)	-0.330 (0.047)	

Item Parameters from PIRLS 2016 Concurrent Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
R41I04C	0.685 (0.026)	0.470 (0.023)		0.301 (0.038)	-0.301 (0.042)	
R41I05M	1.127 (0.080)	0.164 (0.054)	0.223 (0.025)			
R41I06M	1.260 (0.102)	0.487 (0.047)	0.283 (0.021)			
R41I07C	0.968 (0.030)	0.502 (0.017)		-0.008 (0.030)	0.008 (0.032)	
R41I08M	1.594 (0.099)	-0.579 (0.052)	0.264 (0.030)			
R41I09C	1.152 (0.044)	-0.035 (0.023)				
R41I10M	0.952 (0.079)	0.265 (0.070)	0.244 (0.029)			
R41I11C	0.820 (0.030)	0.471 (0.020)		0.256 (0.033)	-0.256 (0.036)	
R41I12M	1.061 (0.070)	0.155 (0.051)	0.145 (0.025)			
R41I13C	0.804 (0.036)	0.207 (0.030)				
R41I14C	0.801 (0.042)	0.944 (0.040)				
R41I15C	0.755 (0.041)	0.690 (0.037)				
R41T01M	1.085 (0.071)	-0.973 (0.092)	0.242 (0.044)			
R41T02C	0.690 (0.024)	-0.516 (0.028)		0.269 (0.048)	-0.269 (0.037)	
R41T03C	0.943 (0.029)	0.071 (0.017)		0.034 (0.033)	-0.034 (0.030)	
R41T04C	1.239 (0.046)	0.055 (0.021)				
R41T05M	0.756 (0.077)	0.383 (0.095)	0.244 (0.035)			
R41T06C	1.381 (0.052)	-0.579 (0.026)				
R41T07C	0.780 (0.025)	0.622 (0.021)		-0.202 (0.037)	0.202 (0.041)	
R41T08C	1.187 (0.044)	0.058 (0.022)				
R41T09M	1.560 (0.103)	0.500 (0.031)	0.182 (0.016)			
R41T10C	1.264 (0.047)	-0.212 (0.023)				
R41T11C	0.859 (0.022)	0.518 (0.014)		-0.426 (0.043)	0.319 (0.050)	0.106 (0.042)
R41T12M	0.999 (0.086)	-0.050 (0.089)	0.346 (0.035)			
R41T13M	1.111 (0.075)	0.432 (0.041)	0.112 (0.020)			
R41T14C	0.533 (0.030)	0.131 (0.043)				
R41T15M	0.897 (0.091)	0.594 (0.072)	0.235 (0.029)			
R41T16M	1.289 (0.098)	-0.046 (0.064)	0.286 (0.030)			

Appendix 12B: PIRLS Literacy 2016 Item Parameters from Item Calibration

Item Parameters from PIRLS Literacy 2016 Item Calibration

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
Items Shared with PIRLS 2016 (Fixed Item Parameters):						
L21B01C *	0.677 (0.036)	-2.530 (0.108)				
L21B02M *	0.958 (0.067)	-2.353 (0.150)	0.101 (0.072)			
L21B03M *	0.721 (0.048)	-2.406 (0.130)	0.250 (0.000)			
L21B04C *	0.706 (0.032)	-1.260 (0.057)				
L21B05M *	0.883 (0.061)	-1.315 (0.127)	0.223 (0.054)			
L21B06M *	0.440 (0.033)	-1.847 (0.144)	0.250 (0.000)			
L21B07C *	0.745 (0.033)	-1.160 (0.052)				
L21B08C *	0.840 (0.036)	-1.359 (0.053)				
L21B09C *	0.855 (0.036)	-1.140 (0.047)				
L21B10M *	0.660 (0.057)	-0.338 (0.133)	0.192 (0.046)			
L21B11M *	0.979 (0.062)	-1.224 (0.103)	0.186 (0.048)			
L21B12M *	0.738 (0.055)	-2.202 (0.212)	0.127 (0.091)			
L21B13C *	0.542 (0.022)	-0.967 (0.040)		0.693 (0.067)	-0.693 (0.044)	
L21B14M *	1.075 (0.063)	-0.597 (0.066)	0.149 (0.032)			
L21B15C *	0.419 (0.026)	0.318 (0.053)				
L21B16C *	0.435 (0.019)	-0.185 (0.035)		0.092 (0.065)	-0.092 (0.059)	
L21B17C *	0.705 (0.027)	-0.414 (0.027)		0.684 (0.048)	-0.684 (0.035)	
R11F01M *	1.334 (0.049)	-0.627 (0.034)	0.148 (0.018)			
R11F02M *	0.666 (0.038)	-0.848 (0.111)	0.243 (0.037)			
R11F03M *	0.920 (0.039)	-0.666 (0.054)	0.157 (0.024)			
R11F04M *	1.307 (0.053)	-0.831 (0.044)	0.228 (0.023)			
R11F05M *	0.940 (0.045)	-0.255 (0.052)	0.217 (0.022)			
R11F06C *	0.776 (0.023)	-0.152 (0.021)				
R11F07C *	0.503 (0.010)	0.375 (0.018)		-0.896 (0.041)	0.896 (0.043)	
R11F08C *	1.149 (0.029)	-0.328 (0.017)				
R11F09C *	1.011 (0.022)	-0.627 (0.015)		0.074 (0.027)	-0.074 (0.020)	
R11F10C *	0.846 (0.026)	-1.419 (0.039)				
R11F11M *	0.739 (0.045)	0.208 (0.061)	0.192 (0.023)			
R11F12C *	0.618 (0.014)	0.642 (0.017)		-0.471 (0.032)	0.471 (0.036)	
R11F13M *	1.124 (0.054)	-0.199 (0.046)	0.270 (0.021)			

* Items with fixed item parameters estimated in PIRLS 2016 concurrent item calibration.

Item Parameters from PIRLS Literacy 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
L21E01C *	0.796 (0.047)	-3.130 (0.125)				
L21E02C *	0.771 (0.038)	-2.116 (0.083)				
L21E03M *	1.343 (0.084)	-0.521 (0.059)	0.270 (0.029)			
L21E04M *	1.106 (0.074)	-0.378 (0.070)	0.262 (0.031)			
L21E05M *	1.118 (0.076)	-1.568 (0.110)	0.242 (0.052)			
L21E06M *	0.936 (0.081)	-2.465 (0.221)	0.304 (0.097)			
L21E07C *	0.463 (0.018)	-0.810 (0.042)		-0.018 (0.071)	0.018 (0.055)	
L21E08M *	1.023 (0.064)	-0.749 (0.081)	0.201 (0.036)			
L21E09M *	0.620 (0.085)	0.464 (0.145)	0.340 (0.041)			
L21E10C *	0.859 (0.047)	-2.716 (0.101)				
L21E11M *	1.205 (0.074)	-0.817 (0.071)	0.226 (0.035)			
L21E12C *	1.116 (0.057)	-2.136 (0.066)				
L21E13C *	0.528 (0.027)	-0.275 (0.046)				
L21E14C *	0.493 (0.020)	0.186 (0.032)		0.852 (0.053)	-0.852 (0.053)	
L21E15C *	0.795 (0.040)	-2.219 (0.087)				
L21E16C *	0.706 (0.032)	-0.959 (0.051)				
L21E17M *	1.047 (0.065)	-0.467 (0.069)	0.171 (0.032)			
R21K01C *	0.422 (0.013)	-0.891 (0.033)		0.186 (0.052)	-0.186 (0.039)	
R21K02C *	0.807 (0.023)	-0.559 (0.025)				
R21K03M *	1.004 (0.047)	0.081 (0.039)	0.184 (0.018)			
R21K04M *	1.062 (0.096)	0.979 (0.045)	0.391 (0.014)			
R21K05C *	0.969 (0.026)	0.137 (0.017)				
R21K06M *	1.489 (0.067)	0.052 (0.029)	0.281 (0.015)			
R21K07C *	0.682 (0.016)	0.143 (0.015)		0.119 (0.027)	-0.119 (0.027)	
R21K08M *	0.994 (0.052)	0.354 (0.037)	0.197 (0.017)			
R21K09M *	1.203 (0.056)	-0.010 (0.037)	0.246 (0.018)			
R21K10C *	0.785 (0.017)	0.778 (0.015)		-0.397 (0.027)	0.397 (0.031)	
R21K11M *	1.070 (0.056)	0.260 (0.039)	0.240 (0.017)			
R21K12C *	0.576 (0.013)	-0.110 (0.014)		0.446 (0.040)	-0.084 (0.037)	-0.362 (0.034)
Items not Shared with PIRLS 2016 (Estimated Item Parameters):						
L21L01M	0.838 (0.126)	-1.433 (0.168)	0.195 (0.054)			
L21L02M	0.647 (0.138)	-0.588 (0.206)	0.217 (0.056)			
L21L03C	0.516 (0.058)	-1.925 (0.132)				
L21L04C	0.656 (0.066)	-1.881 (0.108)				

* Items with fixed item parameters estimated in PIRLS 2016 concurrent item calibration.

Item Parameters from PIRLS Literacy 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
L21L05C	0.534 (0.043)	-1.394 (0.075)		0.019 (0.143)	-0.019 (0.132)	
L21L06C	0.806 (0.078)	-2.118 (0.101)				
L21L07M	1.031 (0.146)	-1.384 (0.132)	0.182 (0.047)			
L21L08M	0.880 (0.165)	-0.776 (0.160)	0.248 (0.050)			
L21L09C	0.400 (0.054)	-1.017 (0.144)				
L21L10M	1.221 (0.186)	-1.194 (0.119)	0.221 (0.045)			
L21L11C	0.489 (0.064)	-0.125 (0.157)				
L21L12M	1.101 (0.169)	-1.518 (0.149)	0.245 (0.055)			
L21L13C	0.586 (0.038)	-1.530 (0.070)		-1.098 (0.178)	1.098 (0.167)	
L21L14M	1.666 (0.249)	-1.382 (0.096)	0.212 (0.042)			
L21L15C	0.440 (0.044)	-0.886 (0.091)		0.371 (0.159)	-0.371 (0.165)	
L21M01M	0.566 (0.089)	-3.339 (0.368)	0.241 (0.093)			
L21M02M	1.340 (0.228)	-0.245 (0.089)	0.155 (0.027)			
L21M03C	0.403 (0.054)	-3.054 (0.241)				
L21M04C	0.733 (0.070)	-1.325 (0.087)				
L21M05M	1.097 (0.164)	-2.371 (0.169)	0.234 (0.065)			
L21M06C	0.853 (0.079)	-1.889 (0.087)				
L21M07M	1.039 (0.185)	-0.605 (0.126)	0.226 (0.040)			
L21M08C	0.991 (0.094)	-2.201 (0.087)				
L21M09M	0.793 (0.118)	-1.455 (0.168)	0.178 (0.053)			
L21M10C	0.423 (0.035)	-2.004 (0.100)		-0.289 (0.193)	0.289 (0.166)	
L21M11M	1.347 (0.270)	-0.177 (0.104)	0.242 (0.032)			
L21M12C	0.677 (0.067)	-1.633 (0.098)				
L21M13M	1.109 (0.158)	-2.173 (0.150)	0.204 (0.058)			
L21M14C	0.586 (0.047)	-0.850 (0.058)		0.667 (0.133)	0.012 (0.140)	-0.679 (0.150)
L21M15M	1.710 (0.335)	-0.455 (0.095)	0.315 (0.034)			
L21M16C	0.833 (0.079)	-1.393 (0.083)				
L21M17C	1.030 (0.104)	-0.372 (0.077)				
L21M18M	1.162 (0.202)	-0.472 (0.108)	0.180 (0.035)			
L11O01M	0.905 (0.145)	-2.254 (0.226)	0.299 (0.075)			
L11O02C	0.509 (0.058)	-0.799 (0.119)				
L11O03M	1.155 (0.181)	-2.672 (0.180)	0.230 (0.069)			
L11O04M	1.168 (0.171)	-1.652 (0.139)	0.248 (0.052)			
L11O05C	0.896 (0.081)	-1.836 (0.084)				
L11O06C	0.665 (0.064)	-1.491 (0.097)				

Item Parameters from PIRLS Literacy 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
L11O07M	0.628 (0.107)	-1.644 (0.261)	0.228 (0.070)			
L11O08M	0.607 (0.104)	-1.735 (0.277)	0.231 (0.073)			
L11O09C	1.230 (0.110)	-1.684 (0.066)				
L11O10C	1.015 (0.107)	-3.050 (0.117)				
L11O11C	0.720 (0.069)	-1.883 (0.100)				
L11O12M	0.547 (0.087)	-2.109 (0.293)	0.196 (0.073)			
L11O13M	0.991 (0.169)	-0.667 (0.127)	0.200 (0.040)			
L11O14C	0.864 (0.080)	-1.884 (0.089)				
L11O15C	0.678 (0.065)	0.092 (0.086)		0.350 (0.105)	-0.350 (0.148)	
L11O16M	1.216 (0.207)	-0.420 (0.099)	0.176 (0.032)			
L11U01C	0.651 (0.065)	-2.359 (0.125)				
L11U02M	0.978 (0.155)	-2.199 (0.205)	0.291 (0.070)			
L11U03M	1.131 (0.163)	-1.283 (0.121)	0.192 (0.043)			
L11U04C	0.681 (0.066)	-1.759 (0.102)				
L11U05M	1.186 (0.159)	-1.713 (0.121)	0.177 (0.045)			
L11U06M	1.194 (0.167)	-1.405 (0.116)	0.186 (0.043)			
L11U07M	1.092 (0.207)	-0.267 (0.112)	0.198 (0.034)			
L11U08M	1.123 (0.190)	-0.732 (0.118)	0.223 (0.039)			
L11U09M	1.071 (0.176)	-1.305 (0.152)	0.278 (0.051)			
L11U10C	0.534 (0.064)	-0.304 (0.134)				
L11U11C	0.649 (0.052)	-1.701 (0.072)		0.220 (0.130)	-0.220 (0.111)	
L11U12C	0.786 (0.074)	-1.447 (0.087)				
L11U13M	1.386 (0.217)	-1.523 (0.125)	0.268 (0.049)			
L11U14C	0.594 (0.074)	-0.008 (0.141)				
L11A01M	0.958 (0.153)	-2.792 (0.234)	0.259 (0.083)			
L11A02C	0.614 (0.063)	-1.861 (0.110)				
L11A03M	0.759 (0.109)	-1.942 (0.195)	0.182 (0.061)			
L11A04C	0.792 (0.074)	-1.498 (0.084)				
L11A05M	1.228 (0.228)	-0.491 (0.111)	0.252 (0.036)			
L11A06C	0.972 (0.097)	-2.614 (0.103)				
L11A07C	0.811 (0.062)	-2.308 (0.069)		-0.050 (0.130)	0.050 (0.103)	
L11A08M	0.995 (0.140)	-1.418 (0.135)	0.181 (0.047)			
L11A09C	1.208 (0.109)	-1.792 (0.067)				
L11A10C	0.774 (0.074)	-0.964 (0.083)				
L11A11C	0.492 (0.061)	-0.325 (0.144)				

Item Parameters from PIRLS Literacy 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
L11A12CZ	0.659 (0.040)	-1.349 (0.049)		-0.418 (0.149)	0.059 (0.172)	0.358 (0.135)
L11A13CZ	0.625 (0.048)	-1.067 (0.065)		0.007 (0.120)	-0.007 (0.121)	
L11A14M	0.791 (0.126)	-1.094 (0.162)	0.163 (0.050)			
L21C01C	0.744 (0.072)	-1.738 (0.095)				
L21C02M	0.935 (0.135)	-0.997 (0.121)	0.140 (0.039)			
L21C03M	0.882 (0.192)	-0.347 (0.153)	0.255 (0.045)			
L21C04C	1.280 (0.119)	-1.917 (0.069)				
L21C05C	1.261 (0.117)	-1.896 (0.069)				
L21C06M	1.452 (0.223)	-1.213 (0.106)	0.240 (0.043)			
L21C07C	0.952 (0.100)	-2.879 (0.121)				
L21C08M	1.164 (0.203)	-0.679 (0.116)	0.236 (0.040)			
L21C09C	0.416 (0.041)	-1.536 (0.100)		0.910 (0.175)	-0.910 (0.154)	
L21C10M	1.339 (0.184)	-1.772 (0.116)	0.197 (0.049)			
L21C11C	1.095 (0.099)	-1.458 (0.069)				
L21C12C	0.675 (0.046)	-1.850 (0.068)		-0.695 (0.154)	0.695 (0.138)	
L21C13M	0.777 (0.131)	-1.145 (0.182)	0.208 (0.056)			
L21C14C	1.043 (0.099)	-2.016 (0.083)				
L21C15M	1.766 (0.284)	-0.792 (0.081)	0.223 (0.034)			
L21C16C	1.128 (0.103)	-1.246 (0.066)				
L21C17C	0.567 (0.072)	-0.105 (0.139)				
L21H01C	0.820 (0.083)	-2.706 (0.121)				
L21H02M	1.006 (0.154)	-1.771 (0.169)	0.264 (0.060)			
L21H03M	1.152 (0.200)	-0.378 (0.105)	0.180 (0.033)			
L21H04M	1.162 (0.174)	-1.756 (0.145)	0.259 (0.056)			
L21H05M	1.309 (0.196)	-1.756 (0.130)	0.260 (0.054)			
L21H06C	0.761 (0.075)	-2.195 (0.106)				
L21H07M	1.034 (0.158)	-1.118 (0.131)	0.206 (0.045)			
L21H08M	1.472 (0.253)	-0.565 (0.095)	0.241 (0.034)			
L21H09M	1.111 (0.161)	-1.391 (0.129)	0.210 (0.048)			
L21H10M	0.784 (0.140)	-0.873 (0.174)	0.216 (0.052)			
L21H11C	0.558 (0.060)	-1.605 (0.114)				
L21H12C	0.680 (0.068)	-1.299 (0.094)				
L21H13M	1.366 (0.213)	-1.162 (0.112)	0.248 (0.043)			
L21H14C	0.811 (0.077)	-1.702 (0.088)				
L21H15M	1.443 (0.297)	-0.229 (0.104)	0.279 (0.033)			

Item Parameters from PIRLS Literacy 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
L21H16M	1.115 (0.197)	-0.751 (0.130)	0.254 (0.043)			
L11P01M	0.939 (0.163)	-0.866 (0.149)	0.237 (0.046)			
L11P02M	0.938 (0.155)	-1.114 (0.156)	0.244 (0.050)			
L11P03C	0.671 (0.056)	-1.014 (0.066)		0.625 (0.106)	-0.625 (0.110)	
L11P04C	0.845 (0.083)	-2.581 (0.108)				
L11P05M	0.997 (0.146)	-1.540 (0.149)	0.220 (0.052)			
L11P06C	0.803 (0.081)	-2.732 (0.119)				
L11P07C	0.733 (0.071)	-0.856 (0.089)				
L11P08M	1.390 (0.200)	-1.184 (0.099)	0.190 (0.038)			
L11P09M	1.330 (0.205)	-1.436 (0.123)	0.265 (0.047)			
L11P10M	1.325 (0.208)	-1.383 (0.124)	0.273 (0.047)			
L11P11C	0.642 (0.044)	-1.237 (0.064)		-0.431 (0.133)	0.431 (0.131)	
L11P12C	0.565 (0.062)	-0.803 (0.112)				
L11P13C	1.128 (0.102)	-1.167 (0.067)				
L11P14C	0.606 (0.065)	-1.032 (0.105)				

Appendix 12C: ePIRLS 2016 Item Parameters from Item Calibration

Item Parameters from ePIRLS 2016 Item Calibration

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
ePIRLS 2016 Items (Estimated Item Parameters):						
E11B01M	0.583 (0.058)	-1.033 (0.241)	0.245 (0.079)			
E11B02M	0.982 (0.077)	-0.918 (0.130)	0.264 (0.063)			
E11B03C	0.681 (0.042)	1.037 (0.051)				
E11B04C	0.539 (0.035)	-0.719 (0.072)				
E11B05M	1.198 (0.092)	0.600 (0.042)	0.137 (0.022)			
E11B06C	0.683 (0.028)	0.348 (0.023)		0.077 (0.043)	-0.077 (0.043)	
E11B07M	0.788 (0.076)	0.364 (0.094)	0.177 (0.040)			
E11B08C	1.126 (0.050)	-0.447 (0.033)				
E11B09C	1.177 (0.050)	-0.226 (0.028)				
E11B10C	0.601 (0.029)	0.941 (0.034)		0.168 (0.045)	-0.168 (0.057)	
E11B11M	1.217 (0.118)	1.007 (0.042)	0.156 (0.019)			
E11B12C	1.035 (0.049)	0.668 (0.027)				
E11B13C	1.191 (0.051)	-0.041 (0.025)				
E11B14C	0.924 (0.046)	0.709 (0.030)				
E11B15C	0.558 (0.025)	0.025 (0.030)		-0.041 (0.057)	0.041 (0.051)	
E11B16C	0.288 (0.014)	-0.274 (0.047)		-0.830 (0.140)	0.375 (0.135)	0.454 (0.106)
E11B17C	0.437 (0.021)	0.436 (0.028)		0.156 (0.081)	0.379 (0.079)	-0.535 (0.078)
E11M01M	1.302 (0.125)	1.040 (0.040)	0.158 (0.018)			
E11M02C	0.805 (0.049)	-1.633 (0.095)				
E11M03C	0.616 (0.036)	-0.470 (0.054)				
E11M04C	1.111 (0.048)	-0.035 (0.026)				
E11M05M	1.493 (0.130)	-0.438 (0.089)	0.458 (0.044)			
E11M06M	0.834 (0.088)	0.213 (0.116)	0.277 (0.046)			
E11M07M	1.300 (0.111)	0.806 (0.039)	0.165 (0.020)			
E11M08C	1.027 (0.047)	-0.448 (0.036)				
E11M09C	0.598 (0.035)	-0.050 (0.044)				
E11M10M	1.349 (0.090)	-0.124 (0.057)	0.200 (0.034)			
E11M11C	0.534 (0.027)	0.779 (0.034)		0.462 (0.049)	-0.462 (0.059)	
E11M12M	1.229 (0.104)	0.195 (0.068)	0.298 (0.034)			
E11M13C	0.900 (0.049)	1.101 (0.042)				

Item Parameters from ePIRLS 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
E11M14C	0.616 (0.020)	0.530 (0.019)		-0.450 (0.059)	0.651 (0.062)	-0.200 (0.054)
E11M15C	0.967 (0.048)	-0.702 (0.046)				
E11M16C	0.753 (0.023)	0.164 (0.021)		-0.567 (0.048)	0.567 (0.046)	
E11M17C	0.432 (0.019)	0.526 (0.034)		-0.628 (0.071)	0.628 (0.074)	
E11M18C	0.902 (0.045)	0.470 (0.029)				
E11M19M	1.376 (0.116)	0.658 (0.043)	0.190 (0.024)			
E11M20C	0.818 (0.050)	1.169 (0.049)				
E11R01M	0.829 (0.068)	-0.314 (0.115)	0.203 (0.050)			
E11R02C	0.588 (0.037)	-1.276 (0.094)				
E11R03C	0.493 (0.024)	0.179 (0.032)		0.396 (0.058)	-0.396 (0.055)	
E11R04M	1.662 (0.137)	0.861 (0.031)	0.187 (0.017)			
E11R05C	0.666 (0.037)	-0.313 (0.046)				
E11R06C	0.718 (0.042)	-0.861 (0.067)				
E11R07C	0.994 (0.033)	0.400 (0.017)		-0.065 (0.032)	0.065 (0.032)	
E11R08C	0.757 (0.041)	0.647 (0.035)				
E11R09C	0.680 (0.037)	0.111 (0.037)				
E11R10M	1.084 (0.097)	0.085 (0.087)	0.316 (0.040)			
E11R11C	0.682 (0.029)	0.437 (0.024)		0.214 (0.042)	-0.214 (0.043)	
E11R12M	1.749 (0.122)	0.048 (0.045)	0.274 (0.029)			
E11R13M	0.756 (0.087)	0.130 (0.148)	0.306 (0.053)			
E11R14C	0.829 (0.041)	-0.094 (0.035)				
E11R15C	1.247 (0.053)	0.064 (0.024)				
E11R16C	0.787 (0.042)	0.236 (0.033)				
E11T01M	0.758 (0.099)	0.451 (0.134)	0.315 (0.048)			
E11T02C	1.158 (0.060)	-1.152 (0.053)				
E11T03M	0.865 (0.080)	0.387 (0.081)	0.176 (0.036)			
E11T04M	1.404 (0.093)	-0.439 (0.065)	0.217 (0.040)			
E11T05C	0.863 (0.032)	-0.356 (0.025)		0.085 (0.044)	-0.085 (0.034)	
E11T06C	1.013 (0.045)	0.182 (0.026)				
E11T07M	0.977 (0.131)	1.320 (0.065)	0.175 (0.022)			
E11T08C	0.734 (0.038)	-0.060 (0.037)				
E11T09C	0.934 (0.050)	1.070 (0.040)				
E11T10C	1.102 (0.048)	0.269 (0.024)				
E11T11M	0.946 (0.061)	-0.456 (0.078)	0.120 (0.038)			
E11T12M	0.737 (0.074)	-0.451 (0.180)	0.296 (0.067)			

Item Parameters from ePIRLS 2016 Item Calibration (Continued)

Item	Slope (a_j)	Location (b_j)	Guessing (c_j)	Step 1 (d_{j1})	Step 2 (d_{j2})	Step 3 (d_{j3})
E11T13M	1.468 (0.127)	0.433 (0.051)	0.323 (0.027)			
E11T14C	0.805 (0.043)	0.678 (0.034)				
E11T15M	1.955 (0.121)	0.282 (0.030)	0.166 (0.020)			
E11T16M	1.270 (0.095)	-0.502 (0.086)	0.275 (0.049)			
E11T17C	1.082 (0.048)	0.169 (0.025)				
E11T18C	0.975 (0.049)	0.732 (0.030)				
E11Z01M	0.969 (0.098)	-0.248 (0.136)	0.411 (0.052)			
E11Z02C	0.511 (0.023)	0.260 (0.029)		-0.043 (0.057)	0.043 (0.056)	
E11Z03M	0.816 (0.079)	0.220 (0.102)	0.211 (0.043)			
E11Z04C	1.068 (0.053)	-0.954 (0.049)				
E11Z05M	0.627 (0.068)	-0.082 (0.174)	0.225 (0.060)			
E11Z06C	1.356 (0.055)	0.229 (0.020)				
E11Z07M	1.403 (0.099)	0.081 (0.052)	0.228 (0.030)			
E11Z08M	1.133 (0.100)	0.261 (0.072)	0.284 (0.034)			
E11Z09C	0.620 (0.041)	1.212 (0.064)				
E11Z10M	1.202 (0.090)	0.344 (0.051)	0.175 (0.027)			
E11Z11M	1.238 (0.091)	-0.650 (0.091)	0.272 (0.051)			
E11Z12C	0.784 (0.039)	-0.083 (0.035)				
E11Z13M	1.429 (0.097)	-0.369 (0.064)	0.232 (0.039)			
E11Z14C	0.884 (0.034)	-0.498 (0.028)		0.134 (0.047)	-0.134 (0.033)	
E11Z15M	1.291 (0.101)	0.193 (0.059)	0.255 (0.032)			
E11Z16C	0.880 (0.033)	0.146 (0.020)		0.151 (0.036)	-0.151 (0.033)	
E11Z17C	0.710 (0.042)	0.838 (0.042)				
E11Z18M	1.180 (0.096)	0.145 (0.069)	0.245 (0.036)			
E11Z19C	1.035 (0.040)	0.599 (0.018)		0.338 (0.028)	-0.338 (0.032)	
E11Z20C	0.963 (0.037)	0.072 (0.020)		0.214 (0.036)	-0.214 (0.031)	