Examining the Components of Linking Error of Trend Estimation in PIRLS

Gabrielle M. Stanco, TIMSS & PIRLS International Study Center, Boston College, stancog@bc.edu
Michael O. Martin, TIMSS & PIRLS International Study Center, Boston College, martinas@bc.edu
Ina V. S. Mullis, TIMSS & PIRLS International Study Center, Boston College, mullis@bc.edu

Abstract

TIMSS and PIRLS, as well as other large-scale assessments, measure changes in achievement over time by linking one assessment to the next. This study explores the issue of linking error in estimating trends, or changes in achievement over time. As part of documenting the changes in student achievement from each assessment cycle to the next, the results are reported together with their standard errors. Typically, the standard errors incorporate a component due to sampling variance and a smaller component resulting from the use of plausible values and conditioning (imputation variance). Recently, there has been research investigating the error component due to linking consecutive assessments, which results from updating the item pool from assessment to assessment. This study examines the error components in linking the 2001 and 2006 PIRLS assessments, focusing on the error component due to changes in the item parameters from one assessment to the next. As anticipated, this error component was small, most likely because almost half the items were in common between the two PIRLS assessments.

Keywords: *linking error, international, trend estimation*

Introduction

Policy-makers have become increasingly interested in trend estimates, which provide information about changing patterns of student achievement and enable monitoring the results of educational reforms over time. The Trends in International Mathematics and Science Study (TIMSS) has been measuring trends internationally every four years since 1995, with assessments in 1999, 2001, 2003, 2007, and soon 2011. Similarly, the Progress in International Reading Literacy Study (PIRLS) is conducted every five years, including assessments in 2001, 2006, and soon 2011.

Essentially, to measure changes in student achievement over time, trend assessments operate on a regular cycle, administering achievement items to comparable samples of students every three to five years, or so. To maintain comparability across cycles, it is necessary to have a substantial number of items in common between adjacent cycles. In fact, if all the items were in common (i.e., there were no new items) and items behaved in the same way from assessment to assessment, there would be no linking error. However, for a variety of reasons, including the necessity for public disclosure and ensuring the current relevance of each new assessment cycle, it also is necessary to continually update the item pool by releasing items after each assessment and replacing them with newly developed items. Accordingly, each assessment consists of newly developed items as well as items used in previous assessment cycles. The fact that some items are released and replaced by others means that the assessment changes somewhat from cycle to cycle, which introduces some error in linking from one to the next.

Due to the increased visibility and reliance on trend data, examining the precision of these estimates of student achievement and their standard errors is becoming an important research area. Therefore, this study explores the components of linking error in trend estimates, with a particular focus on the PIRLS large-scale international assessment. As a point of clarification, it should be noted that although this linking error is important to estimate, it only applies when comparing achievement results across cycles. Linking error is not an issue when making comparisons within the same assessment cycle.

Trend Measurement in TIMSS and PIRLS

TIMSS and PIRLS approach trend measurement from the perspective that assessments need to evolve in order to remain relevant in a changing world, where new knowledge is discovered daily, there is exponential growth of information, and societal views are in a state of flux. Beyond that, shifting demographics as well as new education policies and methodological advances make it critical for assessments to “keep up” with the current times. Therefore, TIMSS and PIRLS were designed around the idea that there must be continuity with past procedures, but that every assessment cycle must also adjust to the changing educational contexts of the participating countries. That is, each cycle of TIMSS and PIRLS links back to the previous cycle by using approximately half of the items from previous assessments, while moving forward with new portions of the assessment by introducing new items in accordance with the updated assessment frameworks.

With each assessment cycle, TIMSS and PIRLS estimate student achievement on the test items using item response theory (IRT) scaling in conjunction with conditioning and plausible value methodology. The IRT scaling for a cycle begins with item calibration to determine the properties of the items (item parameters). This involves applying the appropriate IRT model to the student response data to estimate item parameters, including item difficulty, item discrimination, and guessing (for multiple-choice items). More specifically, TIMSS and PIRLS use concurrent calibration to link the assessments from cycle to cycle. This approach is consistent with the philosophy of evolving across time because concurrent calibration allows the item parameters to adjust to the changing contexts from assessment to assessment.

PIRLS 2006 Concurrent Calibration

Figure 1 shows the concurrent calibration model used for PIRLS 2006. The right hand side of the top panel shows that scaling the PIRLS 2001 achievement data resulted in a distribution of student results. That is, there was an item calibration in PIRLS 2001 and using those item parameters to "score" the student responses yielded the PIRLS 2001 achievement results published in the *PIRLS 2001 International Report* (Mullis, Martin, Gonzalez, & Kennedy, 2003). More specifically, as explained in the *PIRLS 2001 Technical Report* (Martin, Mullis, & Kennedy, 2003), the process of estimating achievement used information about each student's responses to the administered items and the student's background characteristics to impute students scores, or plausible values. To quantify any error in the imputation process, PIRLS generated five plausible values for each student and conducted all analyses five times. The average of the five analyses is taken as the best estimate of the statistic in question, and the difference among them reflects the imputation error. The PIRLS achievement scale metric was established based on the 2001 data to have a mean of 500 and a standard deviation of 100.

As Figure 1 also indicates (top left), upon publishing the PIRLS 2001 achievement results, about half the PIRLS 2001 items were released to the public and the other half were kept secure to be used again in PIRLS 2006. Then, the lower panel of Figure 1 shows the concurrent scaling for PIRLS 2006. The 2006 scaling included only countries that participated in both 2001 and 2006. For those 26 countries, their data from the entire PIRLS 2001 assessment was rescaled together with their new data from the PIRLS 2006 assessment, which consisted of "trend items" from PIRLS 2001 and newly developed items. This concurrent scaling approach of using both complete sets of data from the two consecutive assessments maximizes the data available for modeling student achievement, and allows the item parameters to adjust from cycle to cycle, consistent with an ever changing world.

As can be seen from the top panel, the PIRLS 2001 items had a set of item parameters from the original item calibration conducted in 2001 on the 2001 data. With the concurrent calibration approach, item parameters for the trend items are not fixed, but are re-estimated with each cycle. Thus, the PIRLS 2006 concurrent calibration, using all the data from 2001 and 2006 for countries that participated in both cycles, resulted in a second set of item parameters for the 2001 items. The second set of item parameters reflect changes in estimating achievement based on the updated 2006 assessment consisting of half new items, but also including all of the 2001 data. The item parameters estimated in 2006 were the item parameters used to estimate student achievement on the PIRLS 2006 assessment. As part of this process, the two sets of parameters for the trend items were compared item by item, and the shifts typically were minor. More fundamentally, the PIRLS 2001 data were rescored using the newly estimated 2006 parameters and the results were checked against the original PIRLS 2001 achievement results. Again, the differences were typically very small.

[Take in Figure 1 about here]

As shown in the lower panel of Figure 1, there was a small change between the distribution of the PIRLS 2001 data under the original item calibration and the distribution of the PIRLS 2001 data under the concurrent calibration. The difference was due to changes in item parameters from the original item calibration in 2001 (based on 2001 data) to the concurrent calibration of 2006 (based on both 2001 and 2006 data). These changes in item parameters are the result of the evolving nature of the assessment, which includes new items, reflecting current contexts. To begin the process of placing the PIRLS 2006 data on the achievement scale created in PIRLS 2001, a linear transformation was performed to match the 2001 data under the concurrent calibration to the distribution of the 2001 data under the original item calibration. This transformation ensured that the mean and standard deviation of the distribution of 2001 data under the concurrent calibration aligned to the mean and standard deviation of the distribution of 2001 data from the original item calibration.

Next, the same linear transformation was applied to the 2006 data scaled under concurrent calibration. This placed the PIRLS 2006 data on the PIRLS 2001 metric, allowing the change in achievement between PIRLS 2001 and PIRLS 2006, the trend measure, to be estimated. That is, change in achievement was estimated by examining the difference between the two distributions.

For the concurrent calibration to work, it was necessary to have passages and items in common between PIRLS 2001 and PIRLS 2006. These trend items formed the basis for linking PIRLS 2006 to PIRLS 2001. The PIRLS 2001 assessment consisted of eight passages, four literary and four informational, with a total of 98 items (worth 133 score points). In PIRLS 2006, half of the assessment from PIRLS 2001 (four passages and 49 items, worth 66 points) was reassessed for the purpose of trend measurement. Accordingly, the PIRLS 2006 assessment consisted of 10 passages and items,[1] five literary and five informational, with a total of 126 items with 165 score points. Four of these passages (and 49 items) were trend passages from 2001 and the other three literary and three informational passages and item sets were newly developed to reflect the more current environment and context of reading literacy (Martin, Mullis, & Kennedy, 2007). In summary, the 2006 concurrent calibration included all items, 49 items unique to 2001, 49 items common to 2001 and 2006, and 76 items unique to 2006 (Foy, Galia, & Li, 2007).

Linking Error in PIRLS 2006

Based on the TIMSS and PIRLS concurrent calibration approach to trend estimation, this analysis attempts to quantify the effect of updating the item pool between the two PIRLS administrations in 2001 and 2006. Following the concurrent calibration in 2006, each PIRLS 2001 item had two item parameter estimates, one from 2001 and one from the concurrent calibration in 2006, and the linking error was computed as a function of the standard errors of the differences in achievement resulting from these item parameters. First, average achievement for each country was estimated using the plausible values generated from the 2001 data with the 2001 item parameters (same as in *PIRLS 2001 International Report*), and then average achievement for each country was estimated for the same data using the plausible values generated using the 2006 item parameters. Since the data were the same under both conditions, the difference between the two achievement estimates reflects the effect of the changes made to item parameters between 2001 and 2006, and its standard error may be considered an estimate of the linking error (Johnson, 2005).

The variability due to linking was estimated by jackknifing[2] the difference between average achievement on the 2001 data estimated using the item parameters originally calibrated in 2001 and average achievement on the same data using the item parameters concurrently calibrated using the 2001 and 2006 data. For this research, there were five plausible values from estimating the average achievement on the 2001 data with the 2001 parameters, and five plausible values from estimating the 2001 data with the 2006 parameters. The achievement difference was computed by subtracting the first plausible value in the "2001" set from the first plausible values in the "2006" set, subtracting the second "2001" plausible value from the second "2006" plausible value, and, so on, until there were five differences. Then, the average difference was computed as the statistic of interest, and its standard error was estimated using the jackknife procedure. This standard error of the difference is considered to represent the

1 In PIRLS 2006, the assessment was extended to include 10 passages instead of 8 passages as in PIRLS 2001.

2 See Kennedy & Trong (2007) for a description of the jackknife procedure implemented in PIRLS 2006.

variability associated with updating the item pool between the 2001 and 2006 PIRLS assessments.

The linking error for PIRLS was estimated separately for each country so that results may be adjusted accordingly. Twenty-six countries and two Canadian provinces participated in both PIRLS 2001 and PIRLS 2006 and results were estimated for each.

Results

Table 1 presents the average achievement on the PIRLS 2001 data for each country from the two item calibrations, and shows the difference. The average achievement and standard error shown first were those originally reported in PIRLS 2001 (Mullis, Martin, Gonzalez, & Kennedy, 2003). Next are shown the average achievement and standard error based on the same data using the item parameters from the 2006 concurrent calibration. Third, the differences between average achievement based on the original 2001 calibration and average achievement based on the 2006 concurrent calibration are shown. The standard errors of the differences are the linking error estimate for linking PIRLS 2001 and PIRLS 2006.

[Take in Table 1 about here]

The difference between the two achievement estimates is displayed for each participant, along with its standard error. These differences between achievement estimates are very small, mostly less than half a score point. The linking errors also were small, ranging from 0.06 to 2.0 with an international average of 0.2.

These results are examined in terms of their practical impact on trends in reading achievement for each participant in PIRLS 2006. Re-estimating the standard error of the PIRLS 2006 trend measurements by including the standard error associated with updating the item pool from 2001 provides a more precise estimate of the PIRLS 2006 trend statistics.

Table 2 presents average reading achievement on PIRLS 2001 and PIRLS 2006 as it was published in the PIRLS 2006 International Report (Mullis, Martin, Kennedy, & Foy, 2007). This table also shows the average difference between the two cycles, and its standard error, computed without reference to linking error. Relying on these traditional standard error estimates, in PIRLS 2006 14 countries showed statistically significant changes in reading achievement between 2001 and 2006.

[Take in Table 2 about here]

The standard error of the difference without linking error is computed simply as

$St. Error\ of\ the\ Difference = \sqrt{SE_1^2 + SE_2^2}$, where SE_1 is the standard error from PIRLS 2001 and SE_2 is the standard error from PIRLS 2006. To include the linking error estimated in this analysis, the linking error estimates for each participant were combined with the existing standard errors in the exhibit. Specifically, the revised standard error of the difference was calculated using the following formula: $\sqrt{SE_1^2 + SE_2^2 + SE_L^2}$, where SE_1 and SE_2 again are the standard errors from PIRLS 2001 and PIRLS 2006, respectively and SE_L is the standard error of the link.

Results reflecting the revised standard error estimates are presented in the last column of Table 2. Comparing the published standard errors of the difference between the 2001 and 2006 scale scores to those standard errors including the linking error demonstrates that the linking error had little impact on the statistical significance of the PIRLS 2006 trend estimates. In most countries, including the linking error increased the published standard error by 0.1 points. Due to these small changes, the statistical significance of the trend estimates remained the same for all but one country—Lithuania. Ignoring linking error, the difference between PIRLS 2001 and PIRLS 2006 average scale scores for Lithuania was -6.35 with a standard error of 3.1, which was a statistically significant decrease in reading achievement between 2001 and 2006 ($\frac{-6.35}{3.1}$ = -2.048 > z_{crit}(1.96)). However, after including the linking error, the standard error became 3.3, making the change in reading achievement not statistically significantly different from zero ($\frac{-6.35}{3.3}$ = -1.924 < z_{crit}(1.96)).

Despite this one change in significance, most participants' trend estimates were not greatly affected by the inclusion of the linking error. For example, Latvia was the country with the highest linking error estimate (2.0). However, adding the linking error to the traditional standard error estimate ($\sqrt{2.3^2 + 2.3^2} = 3.25$) only increased it by 0.6 ($\sqrt{2.3^2 + 2.3^2 + 2.0^2} = 3.81$) and did not affect the statistical significance of the trend estimate (i.e., the difference remained non-significant).

Related Research

The Programme for International Student Assessment (PISA) has investigated the extent of linking error in the context of its corresponding trend measurement techniques and theoretical frameworks (Organisation for Economic Co-operation and Development, 2009). Similar to the TIMSS and PIRLS approach, PISA measures trends in achievement by linking results across administrations through the use of a subset of items in common between both cycles. In contrast to TIMSS and PIRLS, however, PISA scales the second assessment cycle separately from the first cycle and does not use concurrent calibration. Therefore, linking error in PISA is a function of item drift (differences in item difficulty from one administration to the next) on the common items as well as the number of common items across surveys (Monseur & Berezner, 2007, p. 329). Based on these factors (including properties of the common items, such as clustering and partial credit), PISA estimates a common transformation, taking into account the linking error of the item parameters, which is applied uniformly to all participating students and countries. The linking error reported by PISA for the 2006 mathematics assessment was 1.4, similar to the 1.1, on average, found in this study for PIRLS 2006.

Conclusion and Implications

Results from this analysis indicate that trends in the PIRLS 2006 achievement estimates were not greatly affected by including the linking error component in the standard errors. This is likely a result of the PIRLS 2006 design, which reassessed approximately 50 percent of the assessment from PIRLS 2001 for the purpose of trend measurement (Martin, Mullis, & Kennedy, 2007). PIRLS 2006 also maintained continuity of content by replacing released passages with similar passages in line with the *PIRLS 2006 Assessment Framework and Specifications*. Thus, it was expected that the linking error for PIRLS 2006 would be minimal. However, this analysis focused on the effects of the specific items used in the two assessments, and did not consider the possible impact of using other possible replacement items. In other words, this analysis treated items as fixed objects of interest, rather than samples from a larger, unspecified pool of replacement items. Also, a linking error estimation procedure that accounts for the clustering of items according to passages needs to be investigated in future analyses. However, there is reason (e.g., Arora, 2007) to believe that this clustering would have a limited effect on the error estimation.

References

Arora, A. (2007). *Creating a TIMSS 2003 problem-solving scale and examining the problem-solving achievement of United States eighth-grade students in TIMSS 2003*. Unpublished doctoral dissertation, Boston College, Massachusetts.

Foy, P., Galia, J., & Li, I. (2007). Scaling the PIRLS 2006 reading assessment data. In M. O. Martin, I. V. S. Mullis, & A. M. Kennedy (Eds.), *PIRLS 2006 technical report* (pp. 149-172). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Johnson, E. (2005). *Trend linking error in PIRLS and TIMSS.* Unpublished manuscript.

Kennedy, A. M., & Trong, K. L. (2007). Reporting student achievement in reading. In M. O. Martin, I. V. S. Mullis, & A. M. Kennedy (Eds.), *PIRLS 2006 technical report* (pp. 173-194). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (Eds.). (2003). *PIRLS 2001 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (Eds.). (2007). *PIRLS 2006 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Monseur, C., & Berezner, A. (2007). The computation of equating errors in international surveys in education. *Journal of Applied Measurement*, *8*(3), 323-335.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 international report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Organisation for Economic Co-operation and Development. (2009). *PISA 2006 technical report*. Paris, France: OECD Publishing.

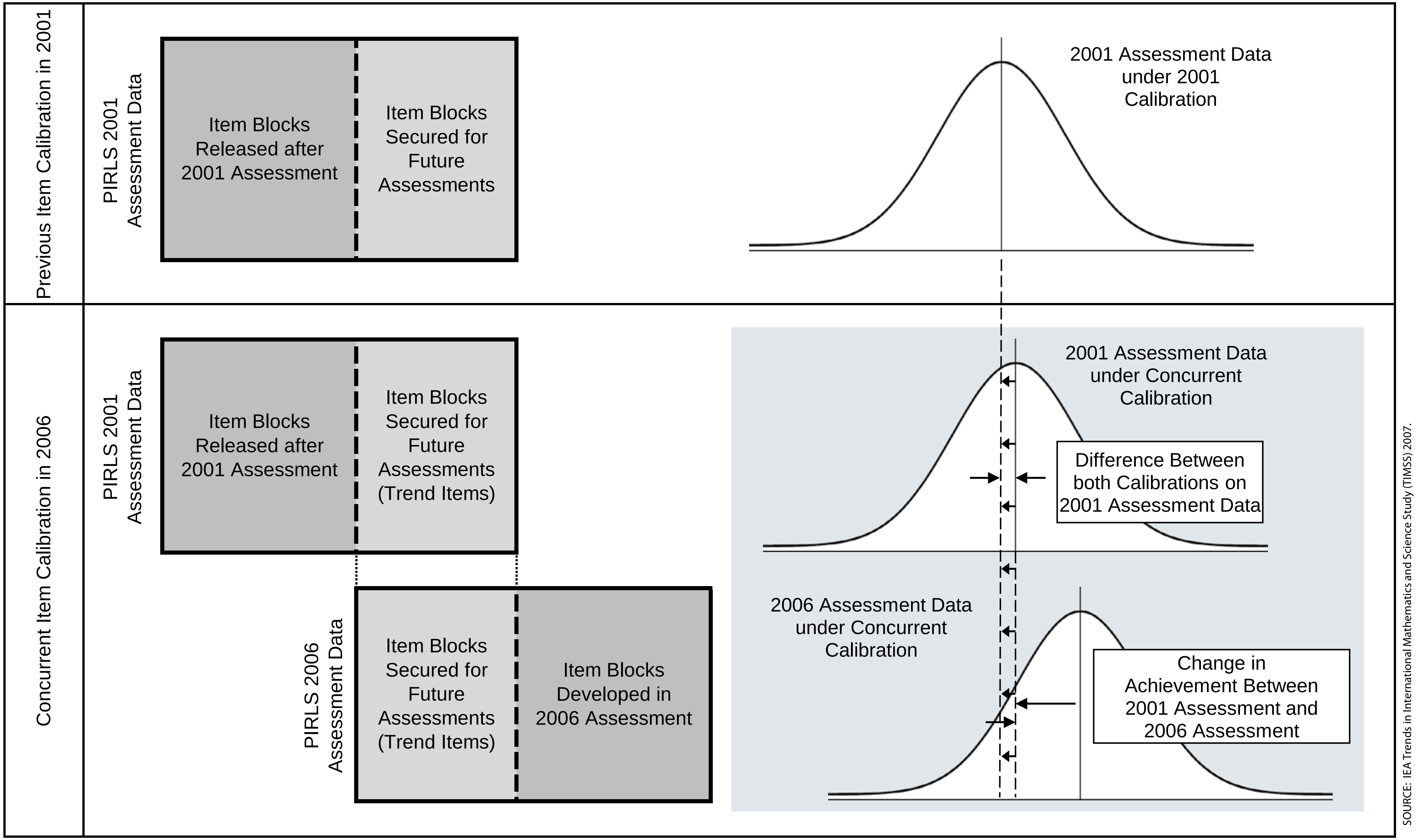

Fig. 1: Concurrent Calibration Model Used for PIRLS 2006

Table 1: Average Achievement and Linking Error Estimates Using PIRLS 2001 Data

Country	2001 Data Estimated from PIRLS 2001 Calibration		2001 Data Estimated from PIRLS 2006 Concurrent Calibration		Difference in Average Achievement	
	Average Achievement	St. Error	Average Achievement	St. Error	Achievement Difference	St. Error of Difference (Linking Error)
Bulgaria	550	3.8	551	3.8	0.3	1.0
England	553	3.4	552	3.3	-0.6	1.4
France	525	2.4	526	2.4	0.4	0.7
Germany	539	1.9	539	1.8	0.0	0.8
Hong Kong SAR	528	3.1	528	3.2	0.3	0.8
Hungary	543	2.2	544	2.1	0.3	1.1
Iceland	512	1.2	512	1.2	-0.7	1.1
Iran, Islamic Rep. of	414	4.2	414	4.4	0.6	1.5
Israel	509	2.8	509	2.8	0.2	1.2
Italy	541	2.4	540	2.4	-0.3	0.8
Latvia	545	2.3	544	2.2	-0.5	2.0
Lithuania	543	2.6	544	2.5	0.2	1.3
Macedonia, Rep. of	442	4.6	442	4.8	0.9	1.1
Moldova, Rep. of	492	4.0	492	4.2	0.0	1.0
Morocco	350	9.6	346	10.0	-3.3	1.5
Netherlands	554	2.5	554	2.7	0.1	1.2
New Zealand	529	3.6	529	3.8	0.2	1.4
Norway	499	2.9	500	2.8	0.7	1.1
Romania	512	4.6	512	4.6	-0.1	0.8
Russian Federation	528	4.4	528	4.2	0.0	1.0
Scotland	528	3.6	528	3.5	0.2	1.0
Singapore	528	5.2	528	5.2	0.1	0.6
Slovak Republic	518	2.8	518	2.8	0.2	1.2
Slovenia	502	2.0	502	1.9	0.3	1.3
Sweden	561	2.2	561	2.3	0.1	1.2
United States	542	3.8	542	3.8	0.3	0.9
International Average	517	0.7	517	0.7	0.00	0.2
Ontario, Canada	548	3.3	548	3.3	-0.1	1.3
Quebec, Canada	537	3.0	538	2.8	0.5	1.1

2001 SE SUM Sqrs	2006 SE	Link SE
14.7970239561	14.2276822416	0.9813477969
11.5207972929	11.0857034304	1.8507697849
5.6049615504	5.9671252729	0.5050092096
3.7436058256	3.3986766025	0.6496198801
9.4782089689	10.1028622500	0.6082908049
4.8370524489	4.4433267264	1.1424830769
1.4376010000	1.5135642729	1.2590635264
17.4905459089	19.2430491561	2.3129238889
8.0356374784	7.7745053584	1.4568731401
5.5302106896	5.8882904964	0.7066419844
5.2172498569	4.9609543824	3.9456657769
6.7010570496	6.3484849444	1.5913822500
21.2522844004	22.8535626916	1.1672641600
15.7333602409	17.3718072025	1.0624661776
93.1199910169	99.6756634129	2.3039596944
6.2372565025	7.0549734544	1.4547978225
12.6943276681	14.4793187289	1.9037928484
8.5405971049	7.5921740521	1.2030141124
21.0612155625	21.6154325776	0.5982713104
19.6451059984	18.0372638209	0.9925140625
12.9682813225	11.9583864481	0.9262330081
26.5805206969	27.1930960900	0.3306480004
8.0979515761	7.6679701921	1.4237978329
3.8637406096	3.6194301504	1.5776616025
4.9203668761	5.4773317369	1.4353956864
14.5675042276	14.6420257201	0.7538580625
11.1570028441	10.6671479236	1.6366340761
8.8370641984	7.7091743716	1.1417349904

Table 2: Trends in Reading Achievement

	Country	PIRLS 2001 Average Scale Score		PIRLS 2006 Average Scale Score		Difference Between PIRLS 2001 and 2006 Scores		
		Average Score	St. Error	Average Score	St. Error	Difference	St. Error Without Linking Error	St. Error Including Linking Error
2a	Russian Federation	528	4.4	565	3.4	37	5.6*	5.6*
	Hong Kong SAR	528	3.1	564	2.4	36	3.9*	4.0*
	Singapore	528	5.2	558	2.9	30	5.9*	5.9*
	Slovenia	502	2.0	522	2.1	20	2.9*	3.2*
	Slovak Republic	518	2.8	531	2.8	13	4.0*	4.1*
	Italy	541	2.4	551	2.9	11	3.8*	3.8*
	Germany	539	1.9	548	2.2	9	2.9*	3.0*
	Moldova, Rep. of	492	4.0	500	3.0	8	5.0	5.1
	Hungary	543	2.2	551	3.0	8	3.7*	3.9*
	Iran, Islamic Rep. of	414	4.2	421	3.1	7	5.2	5.4
2a	*Canada, Ontario*	548	3.3	554	2.8	6	4.4	4.5
2b	Israel	509	2.8	512	3.3	4	4.4	4.5
	New Zealand	529	3.6	532	2.0	3	4.1	4.3
	Macedonia, Rep. of	442	4.6	442	4.1	1	6.2	6.3
†	Scotland	528	3.6	527	2.8	-1	4.6	4.7
‡	Norway	499	2.9	498	2.6	-1	3.9	4.0
	Iceland	512	1.2	511	1.3	-2	1.8	2.1
†2a	United States	542	3.8	540	3.5	-2	5.2	5.2
2a	Bulgaria	550	3.8	547	4.4	-3	5.8	5.9
	France	525	2.4	522	2.1	-4	3.1	3.3
	Latvia	545	2.3	541	2.3	-4	3.3	3.8
	Canada, Quebec	537	3.0	533	2.8	-4	4.1	4.2
	Lithuania	543	2.6	537	1.6	-6	3.1*	3.3
†	Netherlands	554	2.5	547	1.5	-7	2.9*	3.2*
	Sweden	561	2.2	549	2.3	-12	3.2*	3.4*
	England	553	3.4	539	2.6	-13	4.3*	4.5*
	Romania	512	4.6	489	5.0	-22	6.8*	6.8*
	Morocco	350	9.6	323	5.9	-27	11.3*	11.4*

*p<0.05

† Met guidelines for sample participation rates only after replacement schools were included.
‡ Nearly satisfying guidelines for sample participation rates after replacement schools were included.
2a National Defined Population covers less than 95% of National Desired Population.
2b National Defined Population covers less than 80% of National Desired Population.

Trend Note: The primary education systems of the Russian Federation and Slovenia underwent structural changes. Data for Canada, Ontario include only public schools.