The Limits of Measurement: Problems in Estimating Reading Achievement in PIRLS 2006 for Low-performing Countries

Pierre Foy, foypi@bc.edu
Michael O. Martin, martinas@bc.edu
Ina V.S. Mullis, mullis@bc.edu
TIMSS & PIRLS International Study Center
Boston College Abstract

Abstract

This research uses results from PIRLS 2006 to demonstrate how the quality of the measurement based on complex IRT methodology for large scale assessments erodes when the assessment items are too difficult for many of the students. Based on diagnostic information from the PIRLS 2006 IRT scaling, results show that an average percent correct of 30 percent across all the items on the assessment is a useful minimum achievement target, and that countries where fourth grade achievement does not reach this level on the field test should consider administering PIRLS at a higher grade. Such countries also have the option of participating in prePIRLS, IEA's new and less demanding version of PIRLS, which also will be conducted in conjunction with PIRLS in 2011.

Keywords: *Item Percent Correct, Item Response Theory, Large-scale International Assessment, Low-performing countries, Reading Achievement*

Introduction

It is a well-known principle of educational measurement that the difficulty of the items used to assess student achievement should match the ability of the students taking the assessment. In the context of assessing reading comprehension, measurement is most efficient when there is a reasonable match between the reading ability level of the student population being assessed and the difficulty of the assessment passages and items. The greater the mismatch, the more difficult it becomes to achieve reliable measurement. In particular, when the assessment tasks are much too challenging for most of the students, to the extent that many students cannot answer even one question correctly, it is extremely difficult to achieve acceptable measurement quality.

Participating in IEA's PIRLS (Progress in International Reading Literacy Study) is a very resource intensive, high effort endeavor. Further, the ostensible reason for participation is to

obtain high-quality information to use in formulating education policy. However, if the reliability of the students' achievement results has eroded to the point that there is little confidence in the results, participating in PIRLS may not be efficient from a cost/benefit perspective. In this situation, while the experience of participating in a large-scale international assessment is valuable in and of itself, and important background information may be collected, considerable resources and effort will be expended for achievement data that may not be very useful or informative.

IEA is an inclusive organization that encourages participation by as many countries and regional entities as possible. In support of this goal, the TIMSS and PIRLS IRT-based assessment methodology is very robust and provides accurate measurement across a wide range of student abilities. However, like any assessment of student achievement, TIMSS and PIRLS have limits on their measurement capabilities. In particular, as countries with low levels of student achievement consider participating, it is becoming apparent that there needs to be careful consideration of the correspondence between the demands of the assessments and the achievement of the participating students.

The purpose of this research is to use PIRLS 2006 data to demonstrate the degree of erosion in the quality of achievement measurement when the knowledge and skills of the students are considerably below those demanded by the assessment (i.e., the items are too difficult for most students). This research further considers criteria for when the measurement approaches used by PIRLS 2006 have reached their limits, and the reliability of student achievement estimates has decreased to the point that it is no longer meaningful to report the results.

International Performance on PIRLS 2006

When PIRLS was designed originally in 1999, it was targeted at the fourth grade of primary school because this was an important transition point in children's development as readers. In most of the countries intending to participate in PIRLS, fourth grade students essentially had learned how to read and were now reading to learn (Mullis, Kennedy, Martin, & Sainsbury, 2006). Thus, PIRLS is a high-quality, demanding assessment of reading comprehension, designed to measure what accomplished primary school readers can do that have made the transition to reading to learn.

The complete PIRLS 2006 assessment consisted of 10 passages, 5 literary and 5 informational. The literary texts were complete short stories covering a variety of settings, each having one or two main characters and a plot with one or two central events. The informational texts covered a variety of scientific, geographical, biographical, and procedural material and were structured sequentially or by topic. As well as prose, each informational text included organizational and

presentational features such as diagrams, maps, illustrations, photographs, text boxes, lists, or tables. To support the variety of questions necessary to cover the range of comprehension processes assessed, the passages averaged 760 words in length, with a range of from 495 to 872 words (Mullis, Martin, Kennedy, & Foy, 2007).

As a way of describing the overall difficulty of the PIRLS 2006 assessment, the PIRLS 2006 International Report (Mullis, Martin, Kennedy, & Foy, 2007) presented for each country the percentage of students answering each item correctly, averaged across all items in the assessment. In most countries, student achievement was distributed across the full range of the PIRLS assessment, with some students answering most of the items correctly, some students answering few of the items correctly, and most students somewhere in between. Although there were exceptions, in most of the PIRLS 2006 countries there was a reasonable match between the overall difficulty of the assessment and the general level of reading comprehension of the students, and student achievement could be reported with a high degree of confidence.

Most PIRLS 2006 participating countries had an average percent correct in the middle of the percentage range, somewhere from 30 to 70 percent, which seems entirely appropriate for an assessment designed to compare countries across a wide range of achievement levels. The average percent correct across all countries was 54 percent.

There were four countries, however, where students found the PIRLS assessment particularly challenging. These were Kuwait (22% correct), Morocco (21%), Qatar (24%), and South Africa (21%). Although the absolute lower limit in terms of average percent correct is 0 (i.e., no items answered correctly), because about half the items were in multiple choice format and guessing on these was possible, the effective lower limit for the PIRLS 2006 assessment is somewhat higher than this; probably in the region 10-15 percent.[1] With an average percent correct for the country not far above the level a nonreader could approach by chance, it is evident that many students in these countries were unable to engage meaningfully with the PIRLS reading passages and items.

---

[1] Of the 167 score points on the PIRLS 2006 assessment, 64 points came from multiple-choice items and the remainder from constructed response items. Students completely unable to read would likely omit the constructed response items and would receive zero points for these items. However, such students could, in principle, choose an option randomly for each of the multiple choice items, and since each item has four options, could expect to choose correctly on one occasion in four. On average, therefore, such students could expect to answer 16 of the 64 multiple choice items correctly by chance alone. Including the remaining 103 points from constructed response items in the total, such students could expect on average to receive 16 points from the total of 167 points, or approximately 10%. The actual lower limit is probably somewhat higher than this, however, because evidence from the test taking literature suggests that most students who guess have at least some partial information and consequently do not guess completely at random.

In addition to overall reading achievement, PIRLS 2006 reported achievement in terms of two purposes for reading and two reading comprehension processes. The average difficulty of the items addressing reading for literary experience (55%) and reading to acquire and use information (51%) were approximately equal and similar to the overall difficulty level of the assessment. However, undoubtedly reflecting inherent differences in the complexity of the processes themselves, the items addressing the interpreting, integrating, and evaluating comprehension process were more difficult on average than those addressing retrieval and straightforward inferencing (44% correct versus 64% correct).

Not surprisingly, students in the four lowest achieving countries found the interpreting, integrating, and evaluating items to be especially challenging, and on average very few of these items were answered correctly – Kuwait (13%), Morocco (11%), Qatar (14%), and South Africa (14%). Because of this very low performance, student achievement for interpreting, integrating, and evaluating was not reported in these countries in the PIRLS 2006 International Report.

Reporting Student Achievement on PIRLS 2006

As a large scale international assessment, PIRLS was designed to monitor reading comprehension across the entire fourth-grade student population in each country. The 10 reading passages in PIRLS 2006 were distributed among students in each country according to a matrix-sampling design, with each student reading two of the 10 passages. Student achievement on the PIRLS assessment was reported using IRT scaling in conjunction with conditioning and plausible values (Foy, Galia, & Li, 2007). This methodology estimates students' achievement from their responses to the passages that they have read and maps it onto the PIRLS achievement scale,[2] enabling comparisons among student populations and across assessment cycles to measure trends in reading achievement.

The PIRLS scaling methodology is robust, and readily represents the full range of student responses to the PIRLS passages and items on the PIRLS achievement scale in most countries. More specifically, in the large majority of countries where the average percent correct ranges from 30 percent upwards, the PIRLS scaling appears to provide accurate measures of student achievement. However, for countries with lower performance levels, below 30 percent correct, information produced by the scaling process indicates that the PIRLS assessment is not a good

[2] The PIRLS achievement scale was established to have a mean of 500 and standard deviation of 100 corresponding to the mean and standard deviation across the countries that participated in the first cycle of PIRLS in 2001.

fit for many of the students, and that the achievement results produced by the PIRLS scaling may not be reliable indicators of the average level of reading achievement in the country.

Symptoms of Unreliable Measurement in Low Performing Countries

The PIRLS scaling methodology fits a complex model to the data for each country, using each student's pattern of responses to the PIRLS passages and items in conjunction with all available information about the student's background characteristics to estimate the student's likely achievement distribution. Because this distribution is conditional on the student's item responses and background, it is known as the posterior distribution. The student's posterior distribution encapsulates all that we know about the achievement of that student, but the fact that it is a probability distribution (and not a point estimate) indicates that there is still some uncertainty about the student's precise level of achievement.

The posterior distribution is assumed to be approximately normally distributed, conditional on the student's item responses and background characteristics. Therefore, the standard deviation of the posterior distribution, which is a measure of the spread, or dispersion of the distribution, may be taken as a measure of how well the student's achievement has been estimated. The smaller the standard deviation, the narrower will be the possible range of student achievement and the more accurate the estimation. Conversely, the larger the standard deviation, the greater the dispersion will be and the more uncertainty in the achievement estimation.

As the standard deviation of a student's posterior distribution is a measure of how precisely that student's achievement has been estimated, the average posterior standard deviation across all students in a country may be taken as an indicator of how well achievement has been measured in that country, in comparison to other countries. Table 1 presents the average posterior standard deviation for each PIRLS 2006 country for overall reading and for the retrieval and straightforward inferencing and for the interpreting, integrating, and evaluating comprehension processes. Countries are sorted in descending order by average posterior standard deviation on reading overall. It is noteworthy that the four countries where students found the PIRLS 2006 items most challenging – South Africa, Morocco, Kuwait, and Qatar – have the largest average standard deviations of all countries, for overall reading and for each of the two comprehension processes. Not only is student reading achievement low in these countries, it is also not as precisely estimated as in the other countries, and there is more uncertainty involved.

[Take in Table 1 about here]

Given that achievement was measured with least precision overall in the lowest achieving countries, it is useful to examine whether measurement is equally precise all across the achievement spectrum in each country, or whether, for example, it is less precisely estimated among lower achieving students. In order to provide an estimate of student achievement that can be used for analysis and reporting purposes, PIRLS draws five plausible values from the posterior distribution for each student. Because the posterior distribution incorporates all available information about a student's reading proficiency, each plausible value drawn from this distribution is an estimate of the student's achievement. Although each plausible value is an unbiased estimate of the student's achievement, each one includes some error due to the uncertainty of the estimation process. Therefore, group statistics such as means or percentages should be computed by first calculating the statistic separately for each plausible value and then averaging across the five estimates to get the best result. The variability among the five estimates of the statistic reflects the error from the imputation process.

Using the first plausible value[3] as an indicator of student achievement, Figure 1 presents the relationship between student achievement and posterior standard deviation for three countries spanning a range of average achievement on PIRLS 2006, including the Russian Federation (high average achievement – 565 point average), Norway (mid-level average achievement – 498 point average), and Indonesia (mid-to-low average achievement – 405 point average). The figure also shows the relationship across all the PIRLS 2006 countries that had average reading achievement greater than 400, i.e., all countries except the four lowest achieving ones.

[Take in Figure 1 about here]

Figure 1 shows that for each of the three countries and for the aggregate across countries, posterior standard deviations were at their lowest, and consequently measurement at its most precise, at the midpoint of the achievement distribution. As achievement increased above the midpoint, the posterior standard deviation also increased somewhat, as exemplified by the display for the Russian Federation, a country with most students in the upper half of the achievement distribution. The posterior standard deviation also increased as achievement decreased below the midpoint, as may be seen most clearly from the display for Indonesia, a country with most students in the lower half of the distribution. In this display, the standard deviation increased steadily as achievement decreased. The Figure 1 display for the aggregate of all countries with average achievement above 400 clearly shows the u-shaped relationship between achievement and posterior standard deviation. This display also shows that the relationship is somewhat asymmetrical for this set of countries, that is, there were

[3] Plausible values and standard deviations are shown on the logit scale, before transformation to the PIRLS International Scale with its (500, 100) metric.

relatively more students with achievement more than 2 logits below the midpoint than 2 logits above it, and these lower achieving students had generally higher standard deviations. Across this set of countries, the vast majority of students had posterior standard deviations below 0.4 logits.

Figure 2 presents the relationship between reading achievement and posterior standard deviation for the four lowest-achieving countries in PIRLS 2006. As would be expected, in these countries there was a much greater proportion of students in the lower half of the achievement distribution, particularly in the region of the scale below -2 logits, than in the displays shown in Figure 1. It also is very clear that the achievement of the lowest achieving countries was poorly estimated. In Qatar, which had the highest average achievement of the four, about half the students had standard deviations greater than 0.4, and in the other three countries, more than half the students had standard deviations of 0.4 or greater.

[Take in Figure 2 about here]

As the standard deviation of the conditional distribution from which a student's five plausible values are drawn, the posterior standard deviation is one indicator of how precisely achievement has been measured. Another, related indicator is the dispersion among the plausible values themselves, as measured by the standard deviation of each student's five plausible values. Figure 3 provides a perspective on the relationship between achievement and this indicator of measurement precision. It shows how the plausible value standard deviation varies as a function of the first plausible value for the Russian Federation, Norway, and Indonesia, as well as for the aggregate of all PIRLS 2006 countries with average scores above 400. For these countries, which have a reasonably good match between the difficulty of the items and the achievement of the students, there is no clear relationship between achievement and measurement precision, although for Indonesia, which as a country is towards the lower end of the achievement distribution, there is evidence of larger standard deviations among lower achieving students.

[Take in Figure 3 about here]

Figure 4 presents the data on the same relationship for the four lowest achieving countries: Qatar, Kuwait, Morocco, and South Africa. Confirming the message from Figure 2, along with generally low achievement (mostly in the range -4 logits to 0 logits), these countries have many students with relatively large standard deviations. Whereas among countries with average scores above 400, plausible value standard deviations rarely exceeded 0.8 logits (Figure 3), in the lowest achieving countries such standard deviations ranged above 1.2 logits.

[Take in Figure 4 about here]

Why Are Very Low Achievers Badly Estimated?

As a general rule, in order to estimate a student's achievement level and locate the student on an achievement scale it is necessary to present the student with a series of tasks, some of which the student can do successfully and some of which the student is less successful or cannot do. IRT scaling proceeds by examining which items the student answered correctly and which items incorrectly, and based on the difficulty of the items, arrives at an estimate of the proficiency such a student most likely would have. Crucial to finding this proficiency estimate is having both items that the student could answer correctly and items that the student could not answer correctly. Both are necessary; it is the difference between what a student can and cannot do that is at the heart of the proficiency estimation process.

Because the estimation of student proficiency is dependent on evidence of what a student can and cannot do, it is difficult to derive a reliable proficiency estimate when the student can answer very few items correctly and hence provides little evidence of what he or she can do. Fortunately, the marginal estimation procedure used by PIRLS does not rely on accurately estimating each student's proficiency, but rather combines student responses to the items they were administered to estimate the achievement distribution of the entire population. In constructing such population estimates, having a few students with extremely low scores is not a problem, but as the proportion of such students increases, the reliability of the measurement suffers and the threat of bias in the results becomes an important issue.

Bias from Very Low Achievers – The Floor Effect

Constructors of achievement tests have long been aware of the "ceiling effect" in measuring student achievement. A test has a ceiling if the items are too easy for the students taking the test, such that many students answer all of the items correctly. The maximum score on such a test is an underestimate of those students' real achievement, in the sense that they would probably do better on a more difficult test that would allow them to show more of what they are capable of. Perhaps less common in the measurement field is the "floor effect." A test can have a floor effect if it is much too difficult for a student, so that the student cannot answer any of the items correctly. In that situation, the lowest possible score on the test (a "zero" score) may well be an **overestimate** of the student's real achievement, compared to what the results would show on an easier test with items better suited to the student's ability.

In a large scale assessment such as PIRLS, the presence of a floor effect in a country can lead to bias in reporting. If a substantial proportion of students are affected by the floor effect, there will be a substantial proportion of students whose achievement is overestimated, and including such students will lead to overestimation of the achievement distribution of the population as a whole.

Monitoring trends over time is particularly problematic for a country experiencing a floor effect in its assessment. Educators and policy makers can work hard and make real strides in improving education from one assessment cycle to the next, yet not see evidence of this improvement in the assessment results. This can occur because overestimation at the earlier cycle can mask gains from actual improvement at the later cycle. If there are substantial numbers of students with very low scores in the earlier cycle, their achievement may be overestimated[4] and consequently the overall achievement distribution biased upwards. Following four or five years of sustained educational improvements, it could be expected that there would be proportionately fewer very low achievers at the next assessment cycle and therefore fewer students whose achievement has been overestimated. However, because the achievement distribution at the earlier cycle was overestimated to begin with, the difference between the two cycles would not be as big as it would have been if achievement at the earlier time had been estimated accurately. The apparently poor return for all of the effort could be very disheartening to those who worked so hard and could prove a disincentive to further investment and effort.

Limits of Measurement for PIRLS

Experience with PIRLS 2006 has shown that PIRLS at the fourth grade may not be a good match for countries with large proportions of very low achieving students. Although it is clear that the quality of measurement suffers when large proportions of students in a country struggle to answer even the most basic questions on the assessment, it is not clear that there is a particular achievement level below which measurement is not sufficiently reliable. To provide information about low performance on PIRLS, Table 2 lists each country that participated in PIRLS 2006 in descending order by average percent correct across all of the items in the assessment. As well as this global indicator of the relative difficulty of the assessment in each country, the table shows two other indicators of low performance.

[4] Even though these students will have very low scores on the assessment, these low scores overestimate what the students would have scored on an assessment more suited to their achievement level.

The second column in Table 2 shows the percentage of students scoring three score points or less as an indicator of the proportion of students in a country likely to be unable to engage with the passages in the PIRLS assessment. That is, each student reads two passages, which on average offers a maximum of 33 score points. Taking 10 percent of the available points as a realistic estimate for a nonreader, such a student could expect to score 3 points on average on the two passages. As a further perspective on extremely low performance, Table 2 also presents the percentage of students with zero score points (i.e., did not respond correctly to any item), on the assessment overall as well as separately for reading purpose (literary and informational) and comprehension process (retrieval and straightforward inferencing; and interpreting, integrating, and evaluating).

[Take in Table 2 about here]

As mentioned earlier, all but the four lowest performing countries had average percent correct greater than 30 percent, so this may serve as a suitable minimum achievement target in considering whether PIRLS is a good match for the level of achievement in a country. As shown in Table 2, countries with achievement below this level have substantial percentages of students unable to engage with the PIRLS passages (i.e., scored three points or less on the assessment), ranging from 21 percent in Qatar to 41 percent in South Africa. As can be seen from the results of the IRT scaling discussed earlier, it can be a challenge to have reliable measurement with such high proportions of extremely low achievers.

In terms of the more stringent criterion, the percentage of students with zero score points, the results for the overall assessment do not seem alarming at first glance, with South Africa having the highest proportion of such students at 7 percent. However, the figures for the assessment overall do not take into consideration some very large percentages for the interpreting, integrating, and evaluating comprehension process. These range from 23 percent in Qatar to 37 percent in South Africa. It was because of such very low performance that achievement results on this comprehension process were not reported for Qatar, Kuwait, Morocco, and South Africa in PIRLS 2006.

Conclusion

On the basis of the data presented in Table 2, it seems that an average percent correct of 30 percent is a useful minimum achievement target for countries considering administering PIRLS at the fourth grade. In particular, countries planning PIRLS participation for the first time in PIRLS 2011 and who are concerned that their students may not have reached the cognitive competencies of reading comprehension described in the PIRLS 2011 Framework (Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009) should examine their field test results

very carefully. By being based on a series of extended texts dealing with somewhat complicated content, PIRLS is especially challenging for weaker readers. If students cannot persevere through the texts and the field test shows average achievement below the 30 percent correct level, there will not be sufficient response data for reliable estimates in the main assessment. Countries in this situation should consider the likely poor quality of the results and perhaps participate in PIRLS at a higher grade. Also, because IEA is dedicated to widespread participation, it has developed a companion reading comprehension assessment that is based on the PIRLS Framework and uses the same assessment approaches as PIRLS, but is less difficult – called "prePIRLS." Countries can participate in prePIRLS or both prePIRLS and PIRLS depending on their students' capabilities as readers. The goal is to have the measurement match the capabilities of the students to maximize reliability and provide useful, informative results.

References

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.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., Trong, K.L., & Sainsbury, M. (2009). *PIRLS 2011 assessment framework* . Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Kennedy, A.M., Martin, M.O., & Sainsbury, M. (2006). *PIRLS 2006 assessment framework and specifications*. (2nd ed.). 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: IEA's progress in international reading literacy study in primary schools in 40 countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Figure 1: Standard Deviation of Students' Posterior Distribution as a Function of Student Reading Achievement in PIRLS 2006

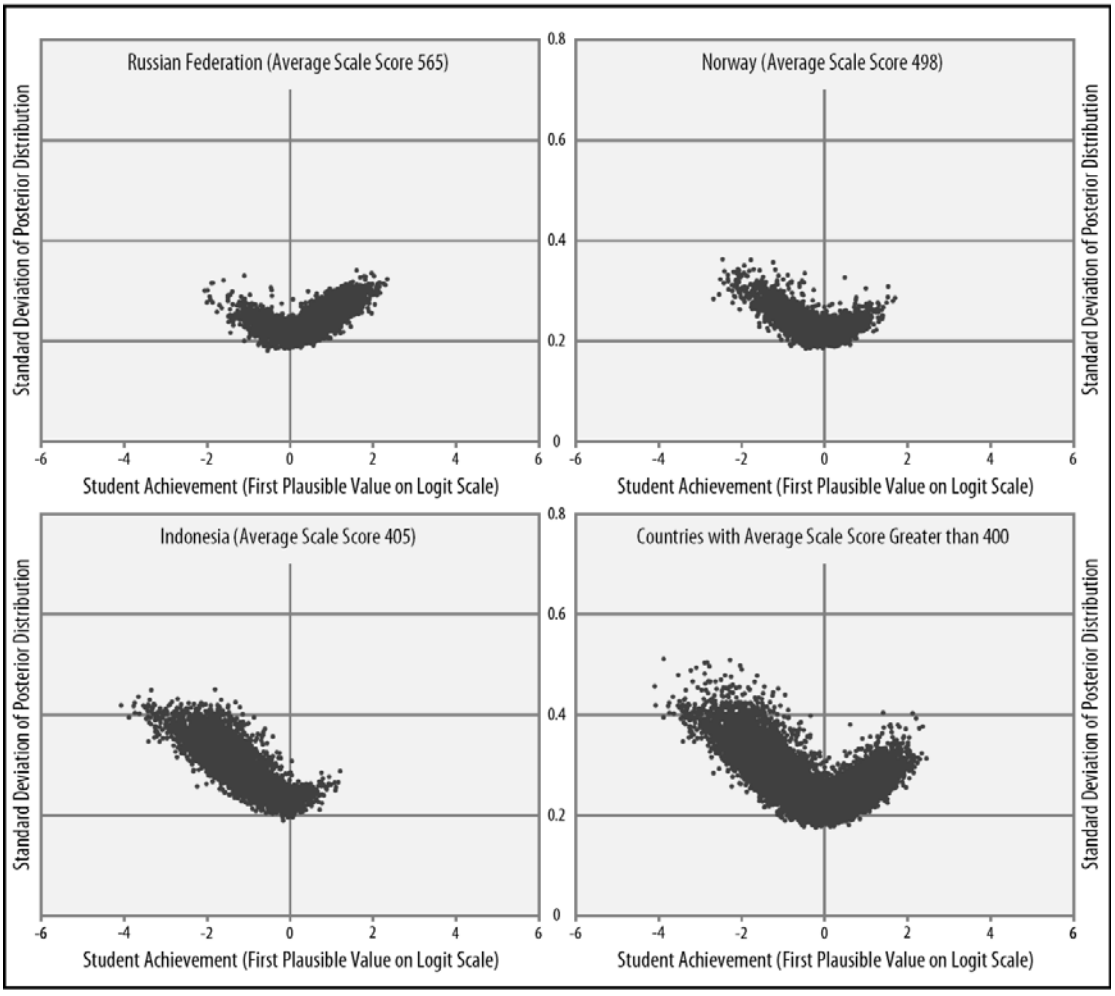

Figure 2: Standard Deviation of Students' Posterior Distribution as a Function of Student Reading Achievement in PIRLS 2006 – Four Lowest Achieving Countries

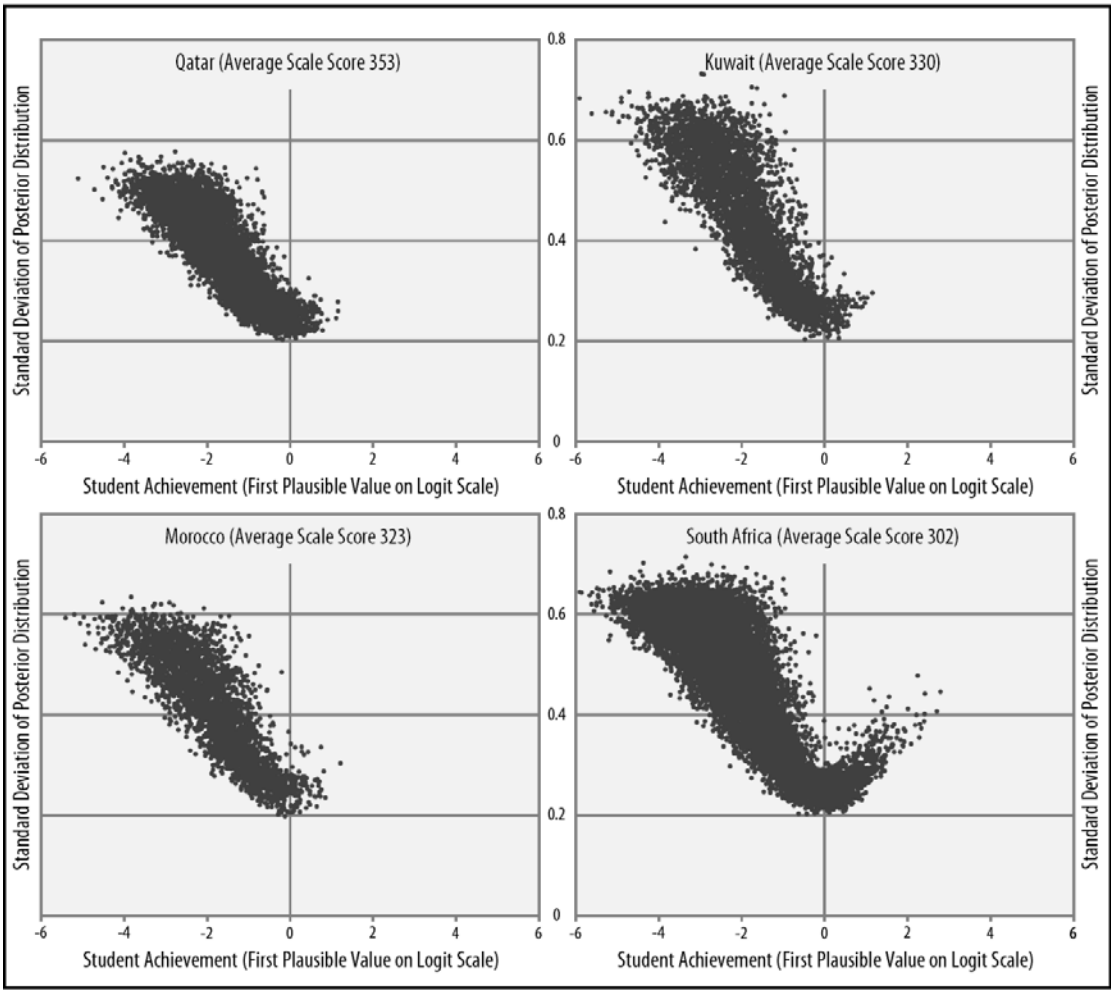

Figure 3: Standard Deviation of Students' Plausible Values as a Function of Student Reading Achievement in PIRLS 2006

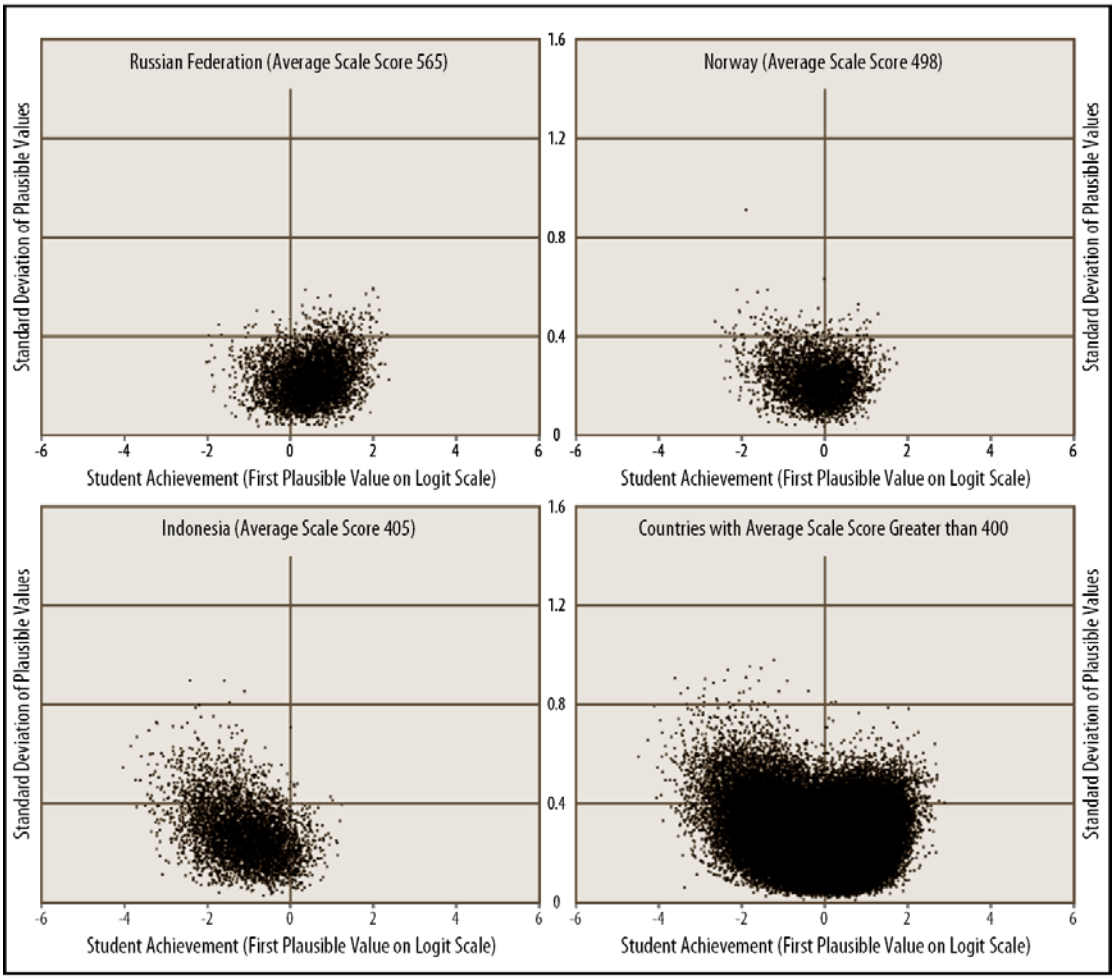

Figure 4: Standard Deviation of Students' Plausible Values as a Function of Student Reading Achievement in PIRLS 2006 – Four Lowest Achieving Countries

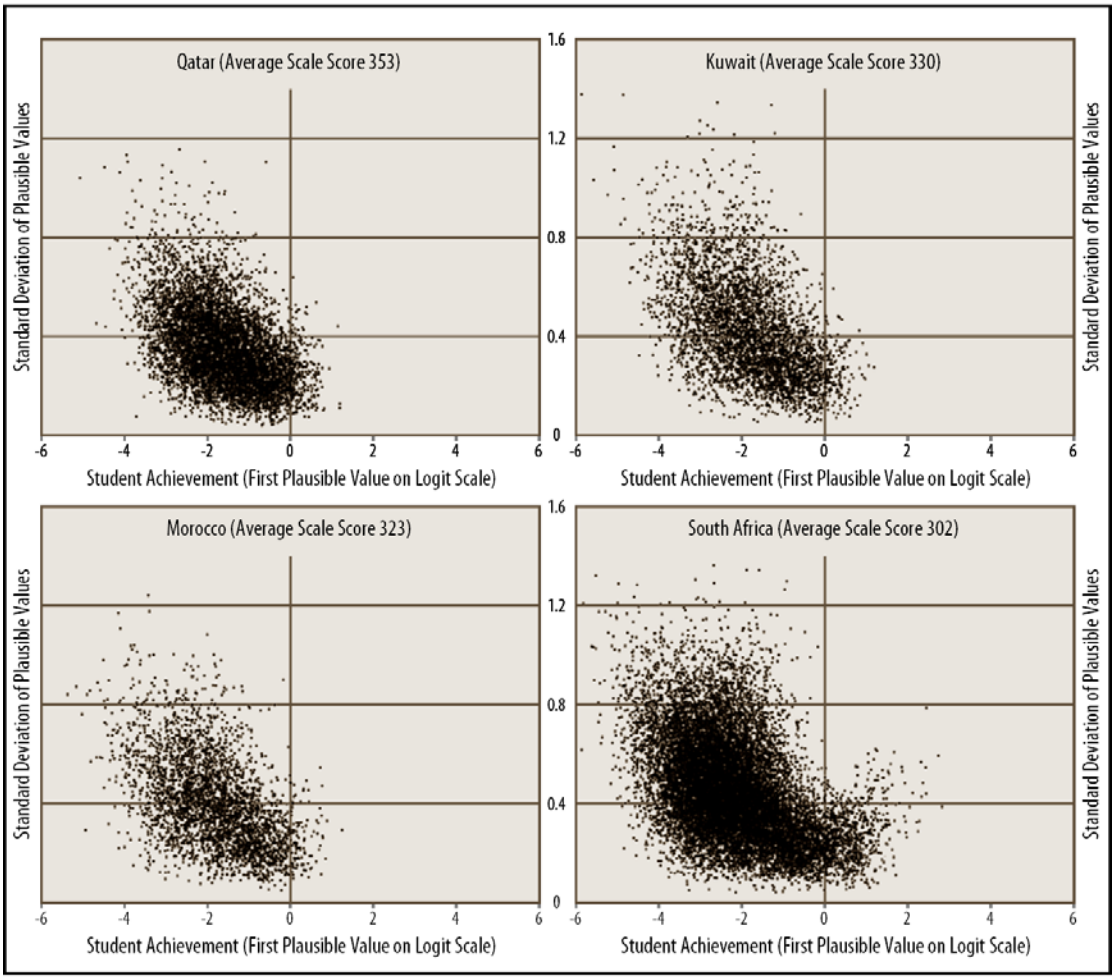

Table 1: Average Posterior Standard Deviation for PIRLS 2006 Participating Countries

Country	Average Standard Deviation of Posterior Distribution		
	Overall Reading	Retrieving and Straightforward Inferencing	Interpreting, Integrating and Evaluating
South Africa	0.47	0.50	0.52
Kuwait	0.45	0.48	0.52
Morocco	0.42	0.42	0.53
Qatar	0.37	0.41	0.40
Iran, Islamic Rep. of	0.31	0.35	0.34
Trinidad and Tobago	0.30	0.31	0.31
Indonesia	0.29	0.32	0.35
Macedonia, Rep. of	0.29	0.30	0.30
Israel	0.27	0.29	0.29
England	0.26	0.25	0.24
Romania	0.26	0.28	0.28
Bulgaria	0.26	0.27	0.27
New Zealand	0.26	0.27	0.27
Scotland	0.25	0.26	0.25
Georgia	0.25	0.27	0.30
United States	0.25	0.25	0.22
Moldova, Rep. of	0.25	0.28	0.28
Singapore	0.24	0.23	0.20
Poland	0.24	0.28	0.26
Slovak Republic	0.24	0.25	0.24
Italy	0.24	0.27	0.25
Denmark	0.24	0.31	0.23
Spain	0.24	0.25	0.25
Russian Federation	0.24	0.25	0.24
Iceland	0.24	0.25	0.24
Luxembourg	0.24	0.30	0.25
Slovenia	0.24	0.25	0.24
Belgium (French)	0.24	0.25	0.24
Chinese Taipei	0.23	0.28	0.25
Norway	0.23	0.21	0.20
Hungary	0.23	0.25	0.24
Hong Kong SAR	0.23	0.26	0.26
France	0.23	0.27	0.26
Sweden	0.23	0.22	0.21
Germany	0.23	0.24	0.23
Austria	0.23	0.25	0.24
Latvia	0.22	0.23	0.22
Belgium (Flemish)	0.22	0.21	0.20
Lithuania	0.21	0.21	0.20
Netherlands	0.21	0.19	0.18

Table 2: Average Percent Correct on the PIRLS 2006 Assessment and Percentage of Students with Very Few Items Correct

Country	Average Percent Correct for Overall Reading	Percentage of Students with Three Points or Less on the PIRLS 2006 Assessment	Percentage of Students with All Items Incorrect on the PIRLS 2006 Assessment				
			Overall Reading	Purposes for Reading		Processes of Comprehension	
				Literary Experience	Acquire and Use Information	Retrieving and Straightforward Inferencing	Interpreting, Integrating and Evaluating
South Africa	21	41	7	11	10	10	37
Morocco	21	30	3	7	7	5	34
Kuwait	22	33	4	6	7	7	33
Qatar	24	21	1	4	5	2	23
Indonesia	31	10	1	2	2	1	14
Iran, Islamic Rep. of	35	10	0	2	2	1	13
Trinidad and Tobago	38	11	1	3	2	1	12
Macedonia, Rep. of	40	10	1	2	2	2	12
Georgia	45	3	0	0	1	0	7
Romania	50	5	1	1	1	1	6
Norway	51	3	0	1	1	1	4
Belgium (French)	51	2	0	0	0	0	3
Moldova, Rep. of	52	2	0	0	1	0	2
Iceland	54	2	0	0	0	0	3
Spain	55	2	0	0	1	0	3
Israel	56	5	0	1	1	0	7
Slovenia	57	1	0	0	0	0	2
Poland	57	1	0	0	0	0	2
France	57	1	0	0	0	0	2
Scotland	59	2	0	0	0	0	2
New Zealand	60	2	0	0	0	0	3
Slovak Republic	60	1	0	0	0	0	2
Chinese Taipei	61	1	0	0	0	0	1
Austria	61	1	0	0	0	0	1
Lithuania	61	0	0	0	0	0	1
England	62	2	0	0	0	0	2
United States	62	1	0	0	0	0	1
Latvia	63	0	0	0	0	0	1
Germany	64	1	0	0	0	0	2
Bulgaria	64	2	0	0	0	0	2
Sweden	64	1	0	0	0	0	1
Denmark	64	1	0	0	0	0	2
Belgium (Flemish)	64	0	0	0	0	0	0
Netherlands	64	0	0	0	0	0	0
Hungary	65	1	0	0	0	0	1
Italy	65	0	0	0	0	0	1
Singapore	66	1	0	0	0	0	1
Luxembourg	66	0	0	0	0	0	1
Russian Federation	68	0	0	0	0	0	1
Hong Kong SAR	69	0	0	0	0	0	1