Examining the Behavior of "Reverse Directional" Items in the TIMSS 2011 Context Questionnaire Scales

Martin Hooper, TIMSS & PIRLS International Study Center, martin.hooper@bc.edu
Alka Arora, TIMSS & PIRLS International Study Center, alka.arora@bc.edu
Michael O. Martin, TIMSS & PIRLS International Study Center, michael.martin.3@bc.edu
Ina V.S. Mullis, TIMSS & PIRLS International Study Center, ina.mullis@bc.edu

Abstract

The TIMSS and PIRLS context questionnaire scales are robust cross-country measures of important aspects of the educational context. While the 2011 context questionnaire scales were very successful, this investigation delves into the question of the effectiveness of using reverse directional items in such scales. Through an exploration of extreme response sets, Rasch fit statistics, and confirmatory factor analysis, this study assesses the relationship between reverse directional items and response patterns in data from the fourth grade and eighth grade *Students Confident in Mathematics* scale. The results suggest that extreme response patterns appear to be rare. In addition, the results of the Rasch fit statistics and the confirmatory factor analysis suggest that the reverse directional items have different psychometric properties than the straightforward items. This analysis recommends further research on whether these patterns of fit of the reverse directional items are manifest through the other TIMSS and PIRLS scales, and whether they may be linked to student reading ability.

Keywords: *TIMSS, PIRLS, Context Questionnaire Scales*

Introduction

The TIMSS and PIRLS international assessments of student achievement in mathematics, science and reading are designed to provide high quality data about policies and practices for improving teaching and learning. As an essential component of these studies, the context questionnaires collect information from students, parents, teachers, and principals about the context for student achievement (Mullis, Martin, Kennedy, Trong, & Sainsbury, 2009; Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009). To make the information from the context questionnaires more useful and interpretable, TIMSS and PIRLS created a range of context questionnaire scales using partial-credit Rasch modeling (Masters & Wright, 1997). Altogether, the TIMSS and PIRLS 2011international reports in mathematics, science, and reading included more than 75 context questionnaire scales describing home, school, and classroom contexts for learning (Mullis, Martin, Foy, & Arora, 2012; Martin, Mullis, Foy, & Stanco, 2012; Mullis, Martin, Foy, & Drucker, 2012).

As described in the TIMSS and PIRLS technical documentation (Martin, Mullis, Foy, & Arora, 2012), the context questionnaire scales proved very successful for the TIMSS and PIRLS 2011 assessments, with acceptable Cronbach's Alpha Reliability coefficients and item-scale loadings, and often a good relationship with student achievement. However, despite the fact that the scales exhibited very good psychometric properties in most countries, there were some instances of less than optimal measurement. The TIMSS and PIRLS context questionnaire scales generally consist of a series of statements about an issue or construct with which a respondent (student, teacher, principal, or parent) is asked to indicate degree of agreement, e.g."Agree a lot", "Agree a little", "Disagree a little", and "Disagree a lot". Usually, the statements express a positive orientation towards the underlying construct, but occasionally statements expressing a negative orientation are included to offset any positive response set, whereby respondents always respond positively, regardless of the orientation of the question. For example, in the *Students Confident in Mathematics Scale, Fourth Grade*, the positive orientation statements like "I am good at working out difficult mathematics problems" are offset by negative orientation statements like "I am just not good at mathematics". The negatively oriented items are reverse coded during scale construction. There was some evidence that these "reverse directional" items sometimes do not behave in the same way as the other items, and that simply reverse coding these items is not sufficient to counter the negative orientation.

This paper examines the behavior of reverse directional items in a TIMSS context questionnaire scale. "Reverse directional items", for the purposes of this paper, are defined as items that show a direction that is opposite to most of the other items on the scale. In the TIMSS 2011 context, all of the scales have a majority of items that show a positive affect towards the construct, and a few of the scales also have reverse directional items that communicate negative affect towards the construct. The items that communicate negative affect may or may not have negative wording like "no", "not", or "never".

Background

The technique of interspersing positive and negative questionnaire items has been debated in the survey methodology literature for over fifty years. Cronbach (1942, 1950, as cited in Billiet & McClendon, 2000) brought to light the "response set" pattern in which respondents mark their answers without reading the content of the item. Subsequent research has built upon this to focus on particular response styles, making a distinction between acquiescence, respondents who tend to agree more often than other participants, and extreme response sets, respondents who tend to always agree regardless of the item content (van Herk, Poortinga, & Verhallen, 2004). Both patterns can make response data difficult to analyze by introducing error into the analysis and making it more difficult to disentangle the true relationships between variables or constructs (Podsakoff, McKenzie, Lee, & Podsakoff, 2003).

As a way of offsetting the effect of extreme response sets, it has become common practice to intersperse items of different directionalities on a scale—meaning that on the same scale some items are

phrased one way in relation to the construct and other items are phrased in the reverse direction. This interspersing of the directionality of items on scales mitigates the effect of extreme response set. In other words, after reverse coding of the items, those who give an extreme response like "Agree a lot", independent of the item content will have more balanced scores as the straightforward directional items will be offset by the reverse directional items (Baumgartner & Steenkamp, 2001; De Vaus, 2002).

Despite the long-established practice of interspersing item directionalities in constructing questionnaire scales, a number of studies have drawn attention to difficulties with this approach. Through factor analysis studies, the interspersing of item direction has been called into question because it has been found that reverse coded items load differently from straightforward items (DiStefano & Motl, 2006; Marsh, 1986). In other words, people tend to have different response patterns for straightforward and reverse items. Also, research has raised questions about how reverse directional items work in the cross-country context (Schmitt & Allik, 2005), and recent research on the TIMSS 2007 data (Marsh et al., 2012) and the TIMSS 2011 data (Castle, 2013) has shown that negative affect items tend to load differently than the positive items. With the development of the TIMSS 2015 and PIRLS 2016 context questionnaires underway, an empirical analysis of these issues could provide information about the continued use of reverse directional items.

To examine the behavior of reverse directional items in TIMSS and PIRLS context questionnaire scales and with a view to making recommendations about future scale development, this research takes as a starting point the TIMSS *Students Confident in Mathematics* scale, which was designed to measure students' confidence in their mathematics abilities at fourth and eighth grades. As shown in Exhibit 1, the fourth and eighth grade versions of these scales are very similar, with seven questionnaire items in the fourth grade version and these same seven (with some minor modifications to wording) plus two more questions in the eighth grade version. Analyzing the scale at fourth and eighth grade is important as research has found that reverse coded items are more difficult for young children (Marsh, 1986).

[INSERT EXHIBITS 1 and 2]

This analysis of the behavior of reverse directional items explores three specific issues:

Analysis of Extreme Response Sets. Since the use of reverse directional items is intended to offset the effect of extreme response sets (De Vaus, 2002), an analysis of the prevalence of response sets provides an indication of the extent to which respondents to a set of questionnaire items reply in the same way regardless of the item content. Although research has found that completely fixed response patterns tend to be rare (Hinz, Michalski, Schwarz, & Herzberg, 2007), it was decided to conduct exploratory analysis of the TIMSS *Students Confident in Mathematics* scale to provide an idea of the prevalence of extreme response sets in the TIMSS data and whether the prevalence of the response sets could vary between countries and between grades.

Analysis of Rasch item fit. In constructing the TIMSS and PIRLS context questionnaire scales, Rasch scale scores are estimated separately for each country, based on a common set of international item parameters. This provides a set of Rasch fit statistics for each item on each scale for each country, enabling a comparison of the fit to the underlying Rasch model for the straightforward and reverse coded items across countries. If the reverse directional items are behaving differently in some countries, it is expected that this would be evident in the fit statistics.

Confirmatory Factor Analysis. Factor analysis has found that reverse directional items load differently from straightforward items (DiStefano & Motl, 2006; Marsh, 1986). Accordingly, this study will use confirmatory factor analysis to examine the behavior of these items in terms of fit to an underlying unidimensional model, and how this fit changes when the effects of reverse directional items are explicitly modeled.

Methodology

Data

This research will be conducted using the TIMSS 2011 database (Foy, Arora, & Stanco, 2013) and will utilize the *Students Confident in Mathematics* scale for the fourth grade and eighth grade. The scale was chosen because it includes numerous reverse directional items, represents the same construct at two grade levels, and has a strong relationship with student achievement.

As can be seen in Exhibit 1, items on the *Students Confident in Mathematics Scale, Fourth Grade*, consist of seven statements about confidence in learning mathematics, and students are asked to indicate their degree of agreement with each one. Three of the items, ASBM03B, ASBM03C, and ASBM03G, are considered to be reverse directional items, and are reverse coded for analysis. Exhibit 2 shows the nine items that make up *the Students Confident in Mathematics* scale, Eighth Grade. Items BSBM16B, BSBM16C, BSBM16E, and BSBM16I are reverse directional items.

Analytic Methods

The three facets of this analysis are (a) to gauge the prevalence of extreme content-independent response sets, (b) to compare Rasch fit statistics for straightforward and reverse directional items, and (c) and to conduct confirmatory factor analysis to investigate whether the covariance structure of the reverse directional items is different from the straightforward items.

Analysis of Extreme Response Sets. Following a methodology similar to the one described by Hinz, Michalski, Schwarz, & Herzberg (2007), an analysis was conducted for each scale to gauge the extent of extreme 'Agree a lot' responses in each country at fourth and eighth grades. Before reverse coding the reverse directional items, the response values of each item on the *Students Confident in Mathematics* scales were summed. Respondents with missing values were omitted using listwise deletion. Because

the summation took place before recoding, the percentage of students who "Agree a lot" to all of the items, regardless of directionality, could be examined. The results report the percentage of students who "Agree a lot" with all items. Because the eighth grade scale has two more items than the fourth grade scale, to aid in the analysis a subset of items also was reported that reflect eighth grade responses to the first seven items of the eighth grade scale. The findings from this analysis are intended to provide a flavor of the prevalence of extreme response sets in the confident scales and how grade level may relate to these response patterns.

Analysis of Rasch item fit. The TIMSS and PIRLS 2011 context questionnaire data were scaled using *Conquest 2.0* (Wu, Adams, Wilson, & Haldane, 2007). For each scale, the item parameters for the Rasch partial-credit model were estimated based on the combined data from all of the participating countries. After the item parameters were calibrated, scale scores were estimated for all respondents (students, teachers, principals, or parents), one country at a time. Because individual scale scores were estimated separately for each country, each had different item fit statistics according to how well the data for the country fit the international calibration model. Using the fit statistics provided by *Conquest 2.0* this analysis examined the fit of the reverse directional items when compared to the fit of the straightforward items. The analysis used the infit mean-square statistic, which compares the items' observed responses to the items' expected responses by examining the average of the item's squared residuals (Bond & Fox, 2007). The primary concern is underfit, which implies the degree of randomness in the response pattern (Linacre, 2002).

Confirmatory Factor Analysis. Confirmatory factor analysis was conducted on both scales to identify differences in the covariance structure of the straightforward items and the reverse directional items. The analysis was conducted using Mplus 7 (Muthen & Muthen, 2013). To accommodate the TIMSS stratified cluster sampling design, the COMPLEX setting was used, and the TIMSS HOUSE WEIGHT sampling weight. The Mplus WLSMV estimator was used for the analysis. Missing values were analyzed based on pairwise deletion. Reverse directional items were reverse coded before analysis. To facilitate the multi-country analysis and the data extraction, the *R* program MPlusAutomation was used to conduct the analysis and extract the data (Halquist, 2012).

The confirmatory factor analysis compared a basic *Students Confident in Mathematics* model (Model 1 and Model 3 in Exhibit 3 for fourth and eighth grades, respectively) to a model that allows covariance in the error terms of the reverse directional items (Model 2 and Model 4). In Models 1 and 3 the error terms are assumed to be uncorrelated, whereas Models 2 and 4 are alternative models with covariances allowed between the error terms of the reverse directional items. By allowing covariance of the error terms of these items, Models 2 and 4 control for the effect of the reverse directional items.

The results of the confirmatory factor analyses on the two models are evaluated using the Root Mean-square Error of Approximation fit index (RMSEA, <.08, acceptable fit), the Comparative Fit

Index (CFI, >.95, acceptable fit) and the Tucker-Lewis Index (TLI, >.95, acceptable fit). Given the large sample sizes in TIMSS countries, we can expect the χ^2 test to be too sensitive for use in this analysis.

[INSERT EXHIBIT 3]

Results

Exhibit 4 presents the percentage of students who 'Agree a lot' to all of the straightforward and reverse directional items on the fourth grade and eighth grade *Students Confident in Mathematics* scale. As can be seen in the exhibit, it is rare at both grade levels that a person would 'Agree a lot' to all of the statements. At the fourth grade, 42 of the 49 countries had less than 3% of their population answer 'Agree a lot' to all of the statements, whereas at the eighth grade, no country had as much as 3% of the population answer 'Agree a lot' to all of the items on the scale, and over 75% of the countries had a negligible percentage of students with this extreme response pattern (<0.5%). Extreme response sets were more prevalent in low performing countries. From this analysis three results emerge: (1) Students marking "Agree a lot" to all items is rare at both grade levels; (2) although rare, the pattern of extreme agreement is more pronounced at the fourth grade level than at the eighth grade; (3) there is variability across countries, with extreme agreement more prevalent in lower performing countries.

[INSERT EXHIBIT 4]

Exhibit 5 shows the mean-square infit statistic of the *Students Confident in Mathematics Scale, Fourth Grade* for each country. The mean-square infit statistic has an expected value of 1, with estimates higher than 1 representing underfit and estimates lower than 1 representing overfit (Bond and Fox, 2007). Linacre (2002) proposes as a rule of thumb that infit statistics of 1.5-2.0 are less productive for measurement, and that infit statistics over 2.0 are potentially a threat to the quality of measurement. As can be seen in Exhibit 5, the *Students Confident in Mathematics Scale, Fourth Grade* shows acceptable fit, although there is variability across items and countries. Most of the countries fall within the range of 0.5 to 1.5 for most of the items indicating that there is generally good fit. Exhibit 6 displays the number of countries with mean-square infit statistics over 1.5 on each item. There is a pattern in which the reverse directional items (ASBM03B, ASBM03C, and ASBM03G) tend to have more underfit than the straightforward items. It is notable that countries with underfit on these items tend to be lower performing countries on the assessment.

[INSERT EXHIBIT 5 AND EXHIBIT 6]

Exhibit 7 displays the mean-square infit statistic of the *Students Confident in Mathematics, Eighth Grade*, and Exhibit 8 graphs the number of countries with an infit mean-square statistic over 1.5

on each item. While there is some underfit for the reverse directional items for some of the countries, the pattern appears to be notably different from that of the fourth grade scale. For the reverse directional item BSBM16C "Mathematics is not one of my strengths", there are no countries over the 1.5 threshold. In contrast, BSBM16C "Mathematics is more difficult for me than for my classmates" and BSBM16I "Mathematics is harder for me than any other subject", there are five and four countries, respectively, with infit statistics over 1.5. BSBM16E "Mathematics makes me confused and nervous" has twelve countries with infit statistics over the 1.5 threshold.

[INSERT EXHIBIT 7 AND EXHIBIT 8]

Exhibit 9 displays the fit statistics of the confirmatory factor analyses of the *Students Confident in Mathematics, Fourth Grade* from TIMSS 2011 (Model 1), and Exhibit 10 shows the results of the model that allows for the covariance between the error terms of the reverse directional items (Model 2). As can be seen in Exhibit 9, the fit statistics for Model 1 tend to be outside of the range of what is generally considered to be acceptable fit (RMSEA, <.08; TLI, >.95; CFI, >.95). However, as can be seen in Model 2 in Exhibit 10, when the error terms of the reverse coded items are allowed to covary, the model fit improves dramatically with most countries achieving acceptable fit indices. Indeed, the average improvement across countries on the RMSEA statistic is 0.09, with fit improving by more than 0.05 in 43 of 49 countries and by over 0.1in 12 of the countries. The variability of the improvement is noticeable with the RMSEA fit improving by less than 0.05 in some countries and improving by over 0.12 in other countries.

[INSERT EXHIBIT 9 AND EXHIBIT 10]

The results of the eighth grade analysis are presented in Exhibit 11 and 12. As can be seen in the Exhibit 11, the fit statistics of the original *Students Confident in Mathematics, Eighth Grade* (Model 3) are higher than what is generally considered acceptable fit (RMSEA, <.08; TLI, >.95; CFI, >.95). Exhibit 12 shows Model 4, the model where the error terms of the reverse directional items are allowed to covary. By including the covariance of the reverse directional items in Model 4, the RMSEA improves in almost all countries. Although the improvement was negligible in countries such as Austria, Hungary and Korea, in 18 of the 42 countries the improvement in the RMSEA statistic was more than 0.05. The average RMSEA improvement when the error terms are allowed to covary is 0.045.

[INSERT EXHIBIT 11 AND EXHIBIT 12]

Because the purpose of this analysis is to suggest a strategy for improve the scales for TIMSS 2015 and PIRLS 2016, and since some of the RMSEA fit statistics were above 0.1 in Model 4, further analysis was conducted to determine what may improve the fit statistics of *Students Confident in Mathematics, Eighth Grade*. In many countries the modification indices pointed to covarying the error

terms between BSBM16G “My teacher thinks I can do well in mathematics <programs/classes/lessons> with difficult materials” and BSBM16H “My teacher tells me I am good at mathematics”. As it seems plausible that students may have a similar response patterns to these items and a teacher loading was found by previous exploratory factor analysis on the TIMSS 2011 dataset (Castle, 2013), the analysis was reanalyzed allowing the error terms from these two teacher items to covary as shown in Exhibit 13. As can be seen in Exhibit 14, when the error terms covary, the fit statistics improved dramatically with all countries reaching an RMSEA under 0.1.

[INSERT EXHIBIT 13 AND EXHIBIT 14]

Discussion

Three primary analyses were conducted on the fourth grade and eighth grade *Students Confident in Mathematics* scales in order to shed light on the functioning of the reverse directional items in the context questionnaire scales.

The analysis of extreme response sets investigated the prevalence of students giving all ‘Agree a lot’ responses to straightforward and reverse directional items. Confirming the research of Hinz, Michalski, Schwarz, & Herzberg (2007), this pattern was found to be rare. However, there was some variability across countries, and extreme response sets were found to be less prevalent in the eighth grade version of the scale compared to the fourth grade version. The fact that very few students seem to follow this fixed response pattern provides some reassurance that students are not mechanically responding to the items regardless of their content. Also, the higher incidence of extreme response sets at the fourth grade implies that there could be an inverse relationship between education level and the tendency to ‘Agree a lot’ to all items. Finally, it was notable that the countries where there appears to be a higher prevalence of extreme response sets tend to be lower performing countries.

The analysis of the Rasch fit statistics shows acceptable infit statistics for all of the items in the *Students Confident in Mathematics* scale in the majority of the countries. Although at the fourth grade, there was some evidence of the reverse directional items not fitting the model as well in some lower performing countries. There was a clear pattern of more underfit on reverse directional items when compared with straightforward items, implying that the response pattern for these items is more random than expected.

The results of the confirmatory factor analysis indicate that the reverse directional items appear to have an effect at the fourth grade in all countries and the eighth grade in certain countries. At the fourth grade, when covariance between error terms of reverse directional items was allowed, the model demonstrated a dramatic increase in fit, suggesting that a reverse directional effect may be present. At

the eighth grade, the introduction of covariance between the error terms led to large improvement in model fit in some countries but only a negligible improvement in others. The results of the confirmatory factor analysis suggest that the reverse directional items may indeed behave differently from the other items, and that this contributes to reduced fit in a simple unidimensional model.

Conclusion

Taken together, the results of this study suggest that the inclusion of reverse directional items in the TIMSS context questionnaire scales may introduce patterns of construct irrelevant variance that complicate the measurement model. The primary justification for the utilization of reverse directional items has been to mitigate the effect of extreme response sets on population statistics. However, given the scarcity of extreme response sets found in this preliminary analysis, this approach may not be justified, at least for the *Students Confident in Mathematics* scale.

Another finding of this analysis is that the effect of the reverse directional items varies across grade levels and countries. The evidence from the Rasch fit stastitics and the confirmatory factor analysis supports the conclusion of Marsh (1986) that reverse directional items do not function as well with young children, as the reverse directional items appear to have worse fit at the fourth grade than at the eighth grade. Similarly, the results confirm the findings of Schmitt and Allik (2005) that the reverse directional items have worse fit in some countries. The results hint to a relationship between country-level item fit and country-level performance on the assessment. Further research should examine this relationship taking into account student reading ability.

In summary, this paper examined the psychometric behavior of the reverse directional items in the *Students Confident in Mathematics* and found evidence that these items do not always behave as anticipated in every country, and that simply reverse coding the responses to these items is not sufficient to make them equivalent in measurement terms to items with straightforward directional orientation. Although results based on a single (albeit at two grade levels) cannot be considered definitive and further research encompassing other TIMSS and PIRLS scales is clearly necessary, the results do suggest that it may be timely to reconsider the role of reverse directional items in future scale development.

References

Baumgartner, H. & Steenkamp, J. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143-156.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal, 7*(4), 608-628.

Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human Sciences*. New York: Routledge.

Castle, C. (2013*) Exploring the latent structure of the TIMSS grade 8 student questionnaire: student attitudes towards science*. Unpublished Manuscript.

Cronbach, L. J. (1942). Studies of acquiescence as a factor in the true-false test. *Journal of Educational Psychology*, *33*, 401–415.

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, *10*, 3–31.

De Vaus, D. (2002) *Surveys in Social Research*. Abingdon, Oxforshire: Routledge.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(3), 440-464.

Foy, P., Arora, A., & Stanco, G. M. (Eds.). (2013). *TIMSS 2011 international database and user guide*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Hallquist, M. (2012). *MplusAutomation*. MplusAutomation: Automating Mplus model estimation and Interpretation [computer software].. R package version 2.15-1, URL: http://CRAN.R project.org/package=MplusAutomation.

Hinz, A., Michalski, D., Schwarz, R., & Herzberg, P. Y. (2007) The acquiescence effect in responding to A questionnaire. *Psychosocial Medicine*, *4*, 1-9.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, *16*(2), 878-879.

Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, *22*(1), 37-49.

Marsh, H. W., Abduljabbar, A.S., Abu-Hilar, M., Morin, A. J. S., Abdelfattah, & F., Leung, K. C. (2012). Factorial, convergent, and discriminant validity of TIMSS math and science motivation Measures: A comparison between Arab and Anglo-Saxon countries. *Journal of Educational, Psychology*, *105*(1), 108-128.

Martin, M. O., Mullis, I.V.S., Foy, P., & Arora, A. (2012). Creating and interpreting the TIMSS and PIRLS 2011context questionnaire scales. In M. O. Martin & I. V. S. Mullis (Eds.), *TIMSS and PIRLS methods and procedures*. Retrieved from http://timssandpirls.bc.edu/methods/t-context-q-scales.html

Martin, M. O., Mullis, I.V.S., Foy, P., & Stanco, G. (2012). *TIMSS 2011 international results in science.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Masters, G. N., & Wright, B. D. (1997). The partial credit model. In: W.J. van de Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory*. Berlin: Springer.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Foy, P., & Drucker, K.T. (2012). *PIRLS 2011 international results in reading.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

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

Mullis, I.V.S., Martin, M. O., Ruddock, G. J., O'Sullivan, C.Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Muthén, L. K., & Muthén, B. O. (1998-2013). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Podsakoff, P. M., Mackenzie, S. B., Lee, J., & Podsakoff, N. P. (2003) Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903.

Schmitt, D., & Allik, J. (2005). Simultaneous administration of the Rosenberg self-esteem scale in 53 Nations: Exploring the universal and cultural specific features of global self-esteem. *Journal of Personality and Social Psychology*, *89*(4), 623-642.

Van Herk, H., Poortinga, Y. H., Verhallen, T. M. M. (2004) Response styles in rating scales: Evidence of method bias in data from six countries. *Journal of Cross-Cultural Psychology*, *35*, 346-360.

Wu, M.L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). *Conquest 2.0* [computer software]. Camberwell, Australia: Australian Council for Educational Research

Exhibit 1: Students Confident in Mathematics Scale, Fourth Grade

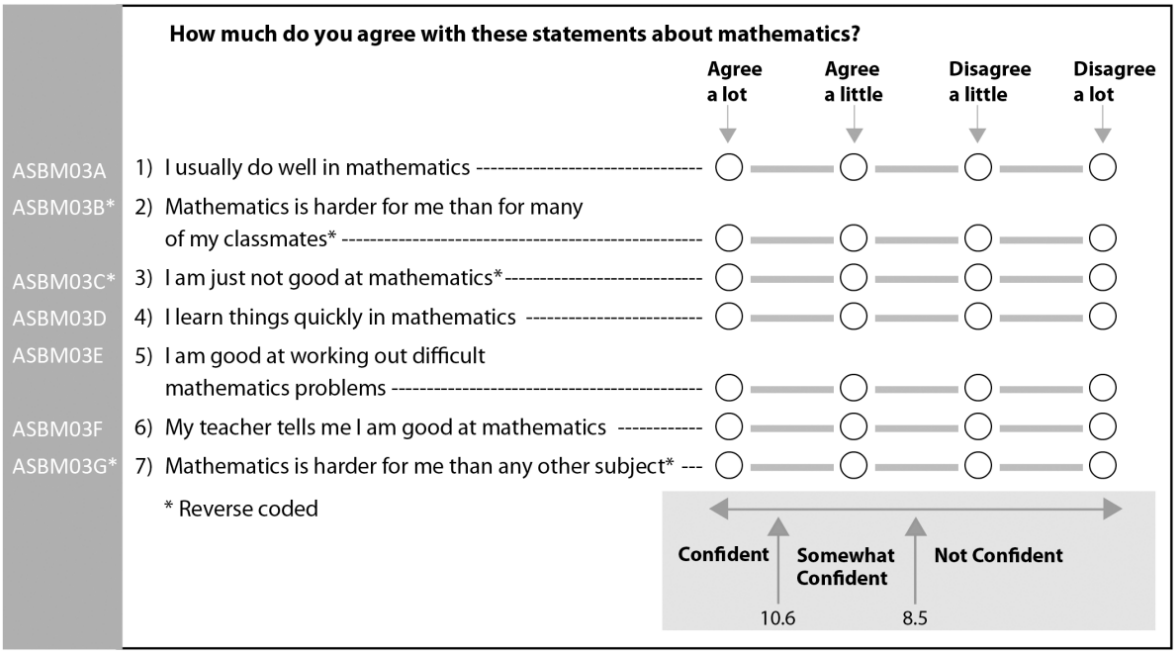


Exhibit 2: Students Confident in Mathematics, Eighth Grade

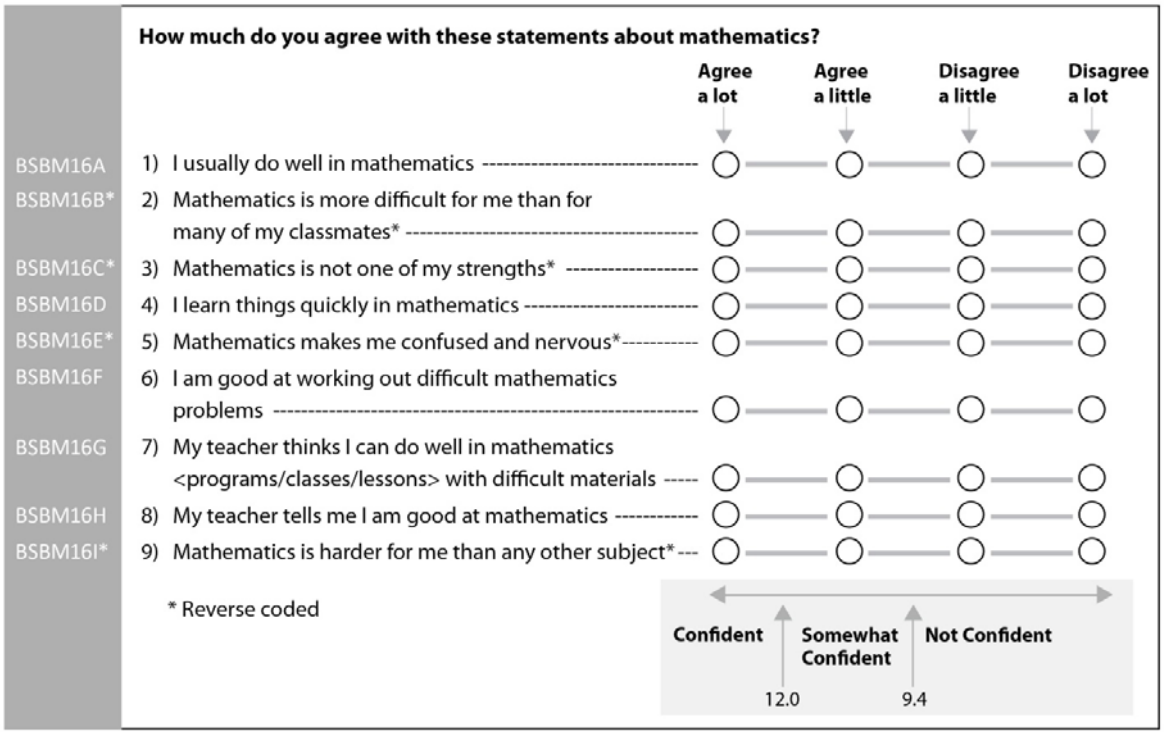


Exhibit 3: Latent variable models for confirmatory factor analysis

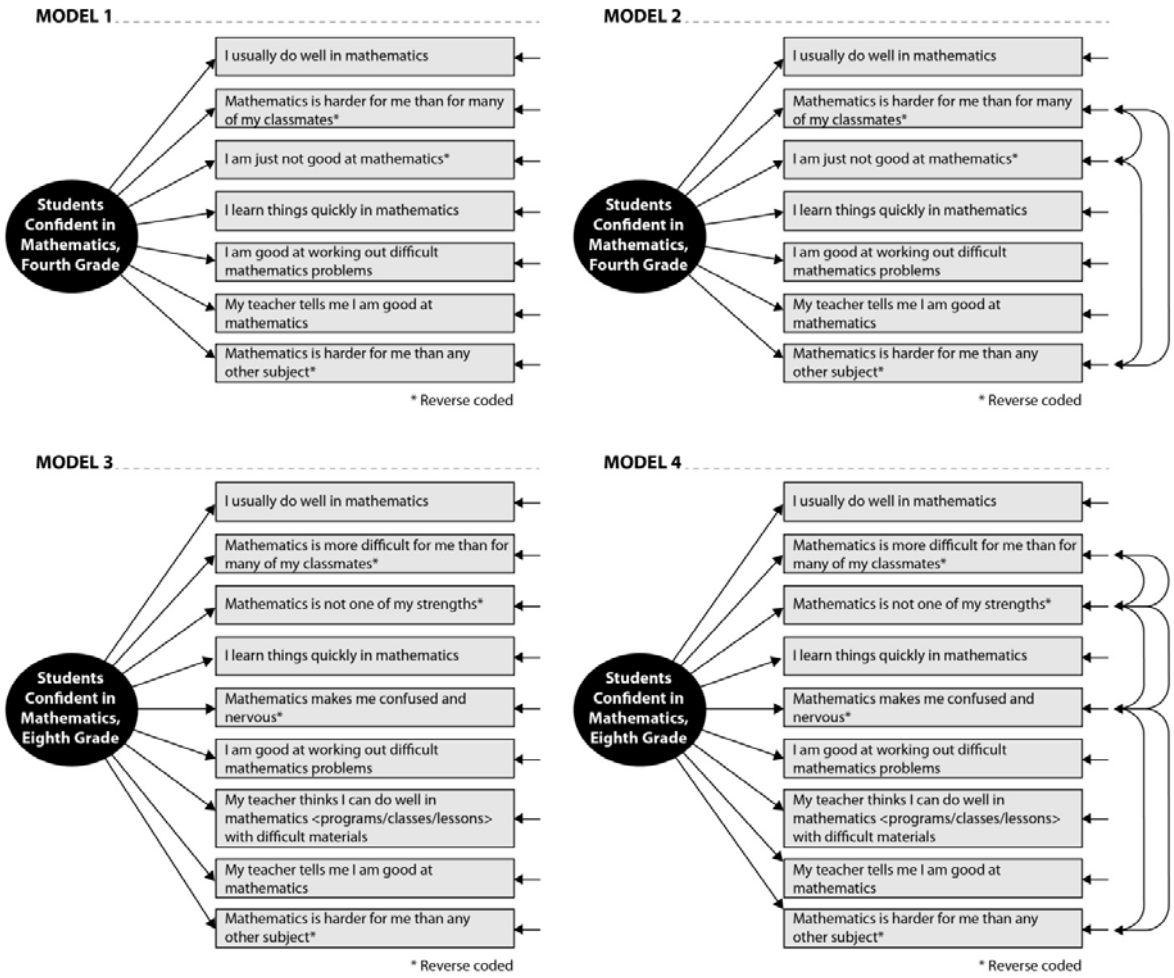


Exhibit 4: Analysis of extreme response sets: percentages of students answering 'Agree a lot' to all items.

Country	Fourth Grade (7 item)	Eighth Grade (9 items)	Eighth Grade (7 items)	Country	Fourth Grade	Eighth Grade (9 items)	Eighth Grade (7 items)
Armenia	1.5	2.1	2.2	**Portugal**	0.5	N/A	N/A
Australia	1.0	0.1	0.1	**Qatar**	5.8	1.3	1.6
Austria	0.5	N/A	N/A	**Romania**	1.3	0.3	0.3
Azerbaijan	6.2	N/A	N/A	**Russian Federation**	0.4	0.0	0.0
Bahrain	2.9	1.3	1.3	**Saudi Arabia**	2.7	0.4	0.4
Belguim (Flemish)*	N/A	N/A	N/A	**Serbia**	0.6	N/A	N/A
Chile	1.9	0.1	0.1	**Singapore**	0.8	0.2	0.2
Chinese Taipei	1.2	0.1	0.1	**Slovak Republic**	0.8	N/A	N/A
Croatia	0.3	N/A	N/A	**Slovenia**	0.7	0.3	0.3
Czech Republic	0.4	N/A	N/A	**Spain**	1.2	N/A	N/A
Denmark	0.1	N/A	N/A	**Sweden**	0.5	0.3	0.3
England	0.3	0.1	0.1	**Syrian Arab Republic**	N/A	1.0	1.2
Finland	0.1	0.1	0.1	**Thailand**	0.7	0.3	0.3
Georgia	2.3	0.3	0.2**	**Tunisia**	1.1	0.3	0.3
Germany	0.4	N/A	N/A	**Turkey**	1.0	0.2	0.2
Ghana	N/A	1.2	1.2	**United Arab Emirates**	3.2	0.7	0.7
Hong Kong SAR	1.2	0.3	0.3	**Ukraine**	N/A	0.1	0.1
Hungary	0.9	0.2	0.2	**United States**	0.7	0.3	0.4
Indonesia	N/A	0.1	0.1	**Yemen**	4.5	N/A	N/A
Iran, Islamic Rep. of	0.0	0.2	0.1				
Ireland	0.5	N/A	N/A				
Israel	N/A	0.3	0.4				
Italy	0.6	0.1	0.1				
Japan	0.0	0.0	0.0				
Jordan	N/A	1.9	2.1				
Kazakhstan	1.6	0.4	0.4				
Korea, Rep. of	0.0	0.0	0.0				
Kuwait	4.6	N/A	N/A				
Lebanon	N/A	0.7	1.0				
Lithuania	0.6	0.2	0.2				
Macedonia, Rep. of	N/A	1.0	1.1				
Malaysia	N/A	0.1	0.1				
Malta	1.6	N/A	N/A				
Morocco	4.8	0.3	0.4				
Netherlands	0.0	N/A	N/A				
New Zealand	1.0	0.1	0.1				
Northern of Ireland	0.7	N/A	N/A				
Norway	0.5	0.2	0.2				
Oman	3.5	0.9	1.1				
PalestinianNat'l Auth.	N/A	0.3	0.4				
Poland	1.1	N/A	N/A				

*Belgium (Flemish) is not included in this analysis as ASBM03E was not administered in this country.

**This analysis used listwise deletion. Georgia's percentage increased from the nine-item eighth grade model to the seven-item ninth grade model. The seven item model had more respondents, as those with missing values on the eighth and ninth items were deleted from the analysis for the nine-item model. As the amount of students who had this response pattern was the same, and the total respondents increased in the seven-item analysis, the percentage of respondent with this response pattern also increased for the seven item model.

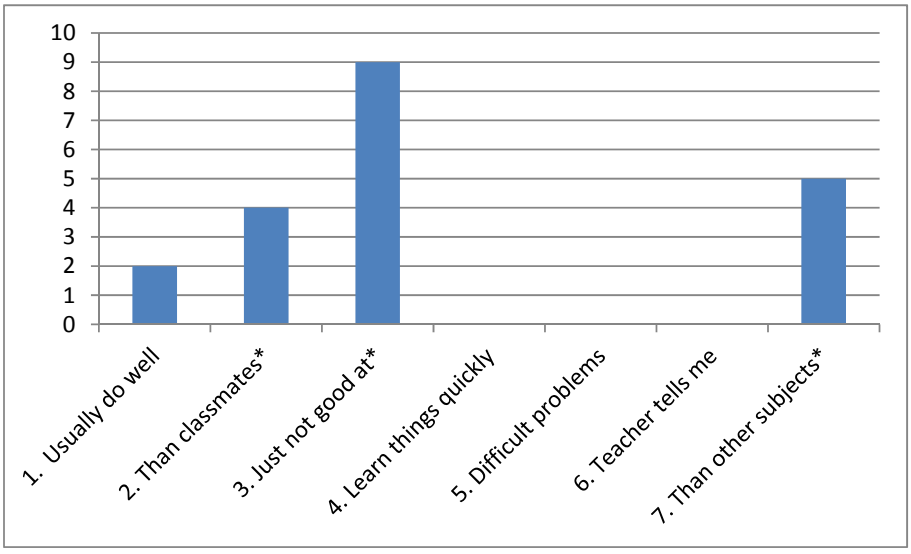
Exhibit 5: Mean-square infit statistic for each item of Students Confident in Mathematics Scale, Fourth Grade

Variable	ASBM03A	ASBM03B*	ASBM03C*	ASBM03D	ASBM03E	ASBM03F	ASBM03G*
Item	I usually do well in Mathematics	Mathematics is harder for me than for many of classmates*	I am just not good at mathematics*	I learn things quickly in mathematics	I am good at working out difficult mathematics problems	My teacher tells me I am good at mathematics	Mathematics is harder for me than other subjects*
Armenia	1.23	1.52	1.35	1.25	1.12	1.31	1.59
Australia	0.79	0.99	0.84	0.94	0.92	1.06	1.04
Austria	0.75	0.93	0.94	0.81	1.17	0.83	0.89
Azerbaijan	1.03	1.33	2.10	0.91	1.17	1.12	1.58
Bahrain	1.25	1.45	1.74	1.06	1.23	1.28	1.31
Belgium (Flemish)**	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Chile	0.91	1.40	1.29	0.98	1.09	1.41	1.27
Chinese Taipei	1.14	1.07	1.00	1.06	0.92	1.12	1.09
Croatia	0.77	0.96	0.87	0.82	1.04	0.80	1.09
Czech Republic	0.83	0.84	0.83	0.87	1.21	0.88	0.81
Denmark	0.61	0.86	0.74	0.78	0.74	0.97	0.88
England	0.69	1.05	0.84	0.94	0.90	1.16	0.97
Finland	0.72	0.90	0.73	0.67	0.97	0.98	0.79
Georgia	0.82	1.43	1.10	0.96	1.31	1.12	1.45
Germany	0.73	0.80	0.77	0.73	0.87	0.81	0.92
Hong Kong SAR	0.81	0.94	0.81	0.80	0.96	1.30	0.90
Hungary	0.86	1.01	0.88	1.00	1.00	0.87	1.11
Iran, Islamic Rep. of	1.09	1.51	1.13	1.05	1.19	1.18	1.42
Ireland	0.65	1.02	0.75	0.77	0.88	1.21	0.79
Italy	0.75	0.94	1.07	0.81	1.03	0.75	0.88
Japan	1.41	0.74	0.64	0.83	0.85	1.41	0.82
Kazakhstan	0.77	0.98	0.94	0.93	0.84	0.85	1.02
Korea, Rep. of	0.91	0.64	0.47	0.79	0.62	1.09	0.72
Kuwait	1.81	1.60	1.55	1.46	1.43	1.50	1.53
Lithuania	0.72	0.99	0.92	0.78	0.83	0.97	1.00
Malta	0.97	1.16	1.11	1.17	1.41	1.14	1.26
Morocco	1.28	1.61	1.99	1.16	1.37	1.32	1.46
Netherlands	0.72	0.93	0.76	0.75	0.82	0.91	0.90
New Zealand	0.82	1.09	0.88	0.98	0.90	1.09	1.07
Northern Ireland	0.79	0.97	0.81	0.91	0.92	1.10	1.03
Norway	0.70	0.95	0.85	0.86	0.92	0.86	0.92
Oman	1.13	1.47	1.69	1.19	1.31	1.25	1.46
Poland	0.71	0.84	1.24	0.78	0.96	1.07	0.91
Portugal	0.60	0.88	0.69	0.72	0.71	0.75	0.92
Qatar	1.22	1.29	1.29	1.12	1.31	1.14	1.31
Romania	0.95	1.47	1.09	0.96	1.14	1.02	1.17
Russian Federation	0.71	0.89	0.83	0.79	0.76	0.85	0.96
Saudi Arabia	1.18	1.37	2.14	1.22	1.34	1.31	1.46
Serbia	0.77	0.88	0.77	0.78	0.83	0.72	0.92
Singapore	0.84	0.88	0.83	0.86	0.91	1.28	0.98
Slovak Republic	0.72	1.09	0.78	0.78	0.82	0.94	0.89
Slovenia	0.72	0.83	0.94	0.77	0.97	0.82	0.88
Spain	0.98	1.24	1.13	1.16	1.50	1.23	1.18
Sweden	0.55	0.63	0.59	0.63	0.74	0.66	0.67
Thailand	0.80	1.12	1.10	1.09	0.91	0.88	1.07
Tunisia	1.89	1.49	1.96	1.07	1.08	1.16	1.73
Turkey	0.91	1.31	1.17	1.06	1.03	1.10	1.45
United Arab Emirates	1.28	1.36	1.63	1.14	1.27	1.11	1.27
United States	0.87	1.03	1.07	1.09	1.02	1.48	1.07
Yemen	1.22	1.49	1.60	1.26	1.28	1.21	1.53

*Reverse coded

**Belgium (Flemish) is not included in this analysis as ASBM03E was not administered in this country.

Exhibit 6: Number of countries with a mean-square infit statistic over 1.5 on each item of the Students Confident in Mathematics Scale, Fourth Grade



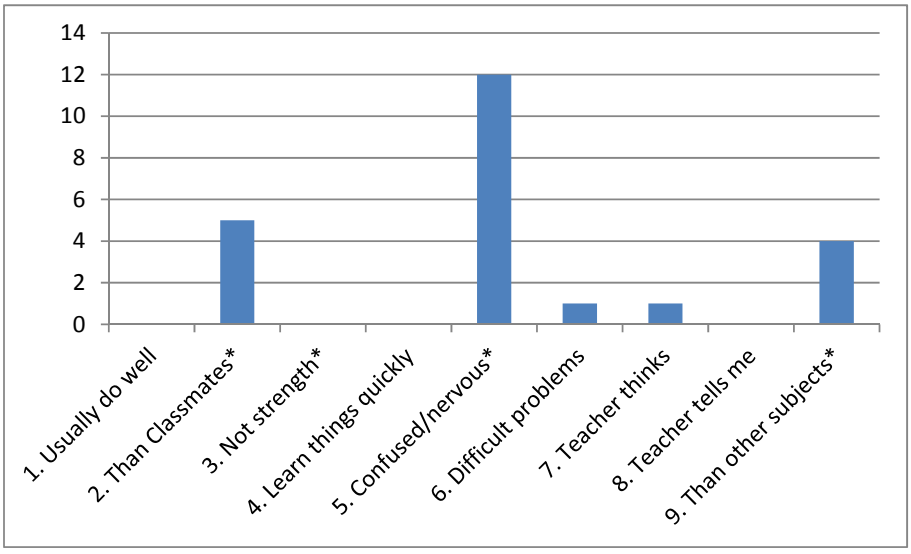
*Reverse Directional Item

Exhibit 7: Mean-square infit statistic for each Item of Students Confident in Mathematics Scale, Eighth Grade

Variable	BSBM16A	BSBM16B*	BSBM16C*	BSBM16D	BSBM16E*	BSBM16F	BSBM16G	BSBM16H	BSBM16I*
Item	I usually do well in mathematics	Mathematics is more difficult for me than for many of my classmates*	Mathematics is not one of my strengths*	I learn things quickly in mathematics	Mathematics makes me confused and nervous*	I am good at working out difficult mathematics problems	My teacher thinks I can do well in mathematics <programs /classes /lessons> with difficult materials	My teacher tells me I am good at mathematics	Mathematics is harder for me than any other subject*
Armenia	0.96	1.24	1.43	1.00	1.52	0.97	1.30	0.95	1.31
Australia	0.64	0.86	0.76	0.69	0.82	0.76	0.87	1.15	0.88
Bahrain	1.13	1.34	1.21	1.05	1.51	1.14	1.40	1.20	1.27
Chile	0.78	1.27	1.11	0.83	1.29	0.89	1.10	1.10	1.35
Chinese Taipei	0.66	0.95	0.83	0.67	1.14	0.64	1.31	0.77	0.94
England	0.62	0.85	0.78	0.77	1.06	0.71	0.74	1.11	0.99
Finland	0.66	0.86	0.66	0.57	0.86	0.77	0.74	0.85	1.18
Georgia	1.02	1.90	1.15	1.00	1.34	1.11	1.34	1.39	1.74
Ghana	1.04	1.46	1.38	1.27	1.65	1.57	1.36	1.22	1.45
Hong Kong SAR	0.72	0.86	0.79	0.70	0.89	0.74	0.91	1.00	0.91
Hungary	0.80	1.26	0.89	0.85	1.32	0.86	1.06	0.79	1.15
Indonesia	0.70	0.59	0.55	0.57	0.60	0.69	0.70	0.59	0.59
Israel	0.80	1.07	1.07	0.93	1.38	0.94	1.35	1.26	1.08
Italy	0.61	0.85	0.62	0.60	0.87	0.74	0.82	0.58	0.80
Japan	1.11	0.89	0.81	0.65	1.22	0.76	1.13	1.06	1.05
Jordan	1.25	1.65	1.41	1.16	1.63	1.19	1.14	1.24	1.45
Kazakhstan	0.64	0.72	0.63	0.74	0.86	0.57	0.70	0.56	0.71
Korea, Rep. of	0.83	0.78	0.58	0.61	0.76	0.51	0.79	0.70	0.72
Lebanon	0.99	1.35	1.31	1.04	1.60	1.27	1.33	0.98	1.20
Lithuania	0.69	1.02	1.02	0.69	1.28	0.74	0.92	1.02	0.95
Malaysia	0.93	1.00	0.97	0.87	0.93	0.81	1.15	1.04	0.97
Macedonia, Rep. of	1.01	1.71	1.42	1.06	1.37	1.12	1.38	1.04	1.26
Morocco	1.28	1.52	1.37	1.12	1.56	1.21	1.39	1.06	1.46
New Zealand	0.61	0.83	0.79	0.73	0.88	0.70	0.78	1.05	0.89
Norway	0.62	0.97	0.85	0.68	0.98	0.73	0.91	0.88	0.90
Oman	1.40	1.33	1.32	1.06	1.46	1.17	1.21	1.32	1.20
Palestinian Nat'l Auth.	1.10	1.50	1.32	1.11	1.57	1.18	1.39	1.29	1.47
Qatar	1.04	1.15	1.15	1.02	1.44	1.08	1.21	1.04	1.15
Romania	1.00	1.28	1.40	1.03	1.66	0.97	1.20	0.97	1.31
Russian Federation	0.74	0.85	0.78	0.82	1.09	0.64	1.03	0.66	0.90
Saudi Arabia	1.16	1.23	1.20	1.08	1.52	1.21	1.22	1.19	1.42
Singapore	0.70	0.75	0.72	0.66	0.87	0.71	0.81	1.00	0.91
Slovenia	0.65	0.73	0.75	0.64	0.87	0.86	0.91	0.60	0.72
Sweden	0.49	0.70	0.64	0.57	0.74	0.64	0.71	0.63	0.72
Syria	1.43	1.62	1.46	1.28	1.66	1.29	1.46	1.27	1.52
Thailand	0.66	0.84	0.69	0.74	0.91	0.79	0.72	0.72	0.80
Tunisia	1.33	1.48	1.31	1.11	1.56	1.09	1.47	1.03	1.52
Turkey	0.83	1.28	1.07	0.92	1.53	1.03	1.54	0.95	1.28
Ukraine	0.67	1.02	0.91	0.85	1.22	0.70	0.79	0.65	1.86
United Arab Emirates	0.95	1.17	1.10	0.91	1.38	1.02	1.17	1.07	1.08
United States	0.76	0.97	0.89	0.81	0.96	0.90	1.08	1.34	0.96

*Reverse coded.

Exhibit 8: Number of countries with a mean-square infit statistic over 1.5 on each item of the Students Confident in Mathematics Scale, Eighth Grade



*Reverse Directional Item

Exhibit 9: Fit statistics for Model 1 factor analysis of Students Confident in Mathematics, Fourth Grade

Country	Observations	X^2	Degrees of Freedom	P-Value	CFI	TLI	RMSEA
Armenia	5106	1547.355	14	0	0.886	0.829	0.146
Australia	6025	1666.633	14	0	0.937	0.905	0.140
Austria	4598	1339.723	14	0	0.947	0.921	0.144
Azerbaijan	4458	1542.280	14	0	0.804	0.706	0.156
Bahrain	4020	914.324	14	0	0.800	0.701	0.126
Belgium (Flemish)*	4807	N/A	N/A	N/A	N/A	N/A	N/A
Chile	5522	2143.893	14	0	0.851	0.776	0.166
Chinese Taipei	4217	2331.462	14	0	0.927	0.890	0.198
Croatia	4569	1658.258	14	0	0.958	0.938	0.160
Czech Republic	4507	1020.345	14	0	0.952	0.927	0.126
Denmark	3829	921.790	14	0	0.959	0.938	0.130
England	3381	669.701	14	0	0.956	0.933	0.118
Finland	4569	1280.343	14	0	0.953	0.930	0.141
Georgia	4663	1821.227	14	0	0.827	0.741	0.166
Germany	3574	1416.610	14	0	0.967	0.951	0.167
Hong Kong SAR	3907	1934.396	14	0	0.928	0.892	0.187
Hungary	5162	1908.557	14	0	0.950	0.925	0.162
Iran, Islamic Rep. of	5716	958.176	14	0	0.876	0.813	0.109
Ireland	4427	979.215	14	0	0.950	0.926	0.125
Italy	4107	1247.199	14	0	0.915	0.873	0.146
Japan	4398	1245.977	14	0	0.954	0.931	0.141
Kazakhstan	4361	1678.789	14	0	0.848	0.772	0.165
Korea, Rep. of	4325	942.627	14	0	0.972	0.958	0.124
Kuwait	4010	954.480	14	0	0.768	0.652	0.129
Lithuania	4627	1698.160	14	0	0.935	0.903	0.161
Malta	3498	1147.251	14	0	0.933	0.900	0.152
Morocco	7191	625.373	14	0	0.627	0.441	0.078
Netherlands	3173	742.584	14	0	0.974	0.960	0.128
New Zealand	5496	1660.181	14	0	0.922	0.882	0.146
Northern Ireland	3510	790.782	14	0	0.954	0.932	0.126
Norway	3035	560.080	14	0	0.928	0.893	0.113
Oman	10209	2592.877	14	0	0.798	0.697	0.134
Poland	4933	1734.108	14	0	0.940	0.910	0.158
Portugal	3997	982.706	14	0	0.939	0.909	0.132
Qatar	3940	2331.456	14	0	0.777	0.666	0.205
Romania	4605	1090.330	14	0	0.896	0.844	0.129
Russian Federation	4450	1958.011	14	0	0.951	0.926	0.177
Saudi Arabia	4337	700.583	14	0	0.742	0.612	0.106
Serbia	4359	1092.238	14	0	0.938	0.907	0.133
Singapore	6298	2772.100	14	0	0.939	0.909	0.177
Slovak Republic	5586	1706.629	14	0	0.910	0.865	0.147
Slovenia	4443	1321.264	14	0	0.927	0.890	0.145
Spain	4078	1011.415	14	0	0.918	0.877	0.132
Sweden	4464	1268.974	14	0	0.951	0.926	0.142
Thailand	4393	1936.889	14	0	0.559	0.339	0.177
Tunisia	4849	336.402	14	0	0.873	0.809	0.069
Turkey	7441	1719.372	14	0	0.849	0.773	0.128
United Arab Emirates	14213	3867.308	14	0	0.830	0.745	0.139
United States	12433	1883.461	14	0	0.952	0.928	0.104
Yemen	7839	346.008	14	0	0.623	0.435	0.055

*Belgium (Flemish) is not included in this analysis as ASBM03E was not administered in this country.

Exhibit 10: Fit statistics for Model 2 factor analysis on Students Confident in Mathematics, Fourth Grade. Model 2 allows covariance between error terms of the reverse directional items.

Country	Students	X^2	Degrees of Freedom	P-Value	CFI	TLI	RMSEA
Armenia	5106	166.269	11	0	0.988	0.978	0.053
Australia	6025	133.250	11	0	0.995	0.991	0.043
Austria	4598	100.384	11	0	0.996	0.993	0.042
Azerbaijan	4458	56.469	11	0	0.994	0.989	0.030
Bahrain	4020	56.732	11	0	0.990	0.981	0.032
Belgium (Flemish)*	4807	N/A	N/A	N/A	N/A	N/A	N/A
Chile	5522	353.210	11	0	0.976	0.954	0.075
Chinese Taipei	4217	277.724	11	0	0.992	0.984	0.076
Croatia	4569	146.665	11	0	0.997	0.993	0.052
Czech Republic	4507	80.579	11	0	0.997	0.994	0.037
Denmark	3829	302.777	11	0	0.987	0.975	0.083
England	3381	203.105	11	0	0.987	0.975	0.072
Finland	4569	96.296	11	0	0.997	0.994	0.041
Georgia	4663	110.341	11	0	0.990	0.982	0.044
Germany	3574	210.004	11	0	0.995	0.991	0.071
Hong Kong SAR	3907	586.317	11	0	0.978	0.959	0.116
Hungary	5162	114.696	11	0	0.997	0.995	0.043
Iran, Islamic Rep. of	5716	61.434	11	0	0.993	0.987	0.028
Ireland	4427	164.909	11	0	0.992	0.985	0.056
Italy	4107	171.174	11	0	0.989	0.979	0.060
Japan	4398	176.131	11	0	0.994	0.988	0.058
Kazakhstan	4361	139.001	11	0	0.988	0.978	0.052
Korea, Rep. of	4325	385.720	11	0	0.989	0.978	0.089
Kuwait	4010	50.567	11	0	0.990	0.981	0.030
Lithuania	4627	270.916	11	0	0.990	0.981	0.071
Malta	3498	119.844	11	0	0.994	0.988	0.053
Morocco	7191	43.000	11	0	0.980	0.963	0.020
Netherlands	3173	129.899	11	0	0.996	0.992	0.058
New Zealand	5496	179.909	11	0	0.992	0.985	0.053
Northern Ireland	3510	203.600	11	0	0.989	0.978	0.071
Norway	3035	89.739	11	0	0.990	0.980	0.049
Oman	10209	68.780	11	0	0.995	0.991	0.023
Poland	4933	353.972	11	0	0.988	0.977	0.080
Portugal	3997	99.448	11	0	0.994	0.989	0.045
Qatar	3940	88.395	11	0	0.993	0.986	0.042
Romania	4605	148.946	11	0	0.987	0.975	0.052
Russian Federation	4450	474.183	11	0	0.988	0.978	0.097
Saudi Arabia	4337	50.884	11	0	0.985	0.971	0.029
Serbia	4359	48.665	11	0	0.998	0.996	0.028
Singapore	6298	659.705	11	0	0.986	0.973	0.097
Slovak Republic	5586	182.908	11	0	0.991	0.983	0.053
Slovenia	4443	164.231	11	0	0.991	0.984	0.056
Spain	4078	185.479	11	0	0.986	0.973	0.062
Sweden	4464	117.547	11	0	0.996	0.992	0.047
Thailand	4393	52.938	11	0	0.990	0.982	0.029
Tunisia	4849	42.496	11	0	0.988	0.976	0.024
Turkey	7441	137.945	11	0	0.989	0.979	0.039
United Arab Emirates	14213	112.854	11	0	0.995	0.991	0.026
United States	12433	489.553	11	0	0.988	0.977	0.059
Yemen	7839	26.313	11	0.0058	0.983	0.967	0.013

*Belguim (Flemish) was not included in this analysis as ASBM03E was not administered in this country

Exhibit 11: Fit statistics for the Model 3 factor analysis of Students Confident in Mathematics Scale, Eighth Grade

Country	Number	X^2	Degrees of Freedom	P-Value	CFI	TLI	RMSEA
Armenia	5767	2966.769	27	0	0.890	0.853	0.137
Australia	7388	2971.906	27	0	0.920	0.893	0.122
Bahrain	4579	1780.461	27	0	0.884	0.845	0.119
Chile	5807	3646.431	27	0	0.896	0.861	0.152
Chinese Taipei	5037	3522.095	27	0	0.968	0.958	0.160
England	3814	2031.513	27	0	0.904	0.872	0.140
Finland	4216	2360.892	27	0	0.969	0.959	0.143
Georgia	4527	2637.147	27	0	0.832	0.775	0.146
Ghana	7214	1949.862	27	0	0.786	0.715	0.099
Hong Kong SAR	3981	2575.097	27	0	0.945	0.926	0.154
Hungary	5165	2297.400	27	0	0.963	0.951	0.128
Indonesia	5761	4340.333	27	0	0.685	0.580	0.167
Iran, Islamic Rep. of	6017	2797.758	27	0	0.883	0.844	0.131
Israel	4645	2074.544	27	0	0.910	0.880	0.128
Italy	3959	2098.820	27	0	0.963	0.951	0.139
Japan	4361	4290.998	27	0	0.900	0.866	0.190
Jordan	7588	2592.093	27	0	0.859	0.812	0.112
Kazakhstan	4377	2001.839	27	0	0.927	0.903	0.129
Korea, Rep. of	5163	4918.475	27	0	0.934	0.912	0.187
Lebanon	3882	1737.669	27	0	0.853	0.805	0.128
Lithuania	4723	2218.826	27	0	0.945	0.927	0.131
Macedonia, Rep. of	3971	3527.270	27	0	0.804	0.738	0.181
Malaysia	5724	4810.153	27	0	0.622	0.495	0.176
Morocco	8837	3580.017	27	0	0.762	0.682	0.122
New Zealand	5191	3364.222	27	0	0.922	0.896	0.154
Norway	3827	1689.881	27	0	0.959	0.945	0.127
Oman	9430	4302.721	27	0	0.715	0.620	0.130
Palestinian Nat'l Auth.	7774	2108.863	27	0	0.786	0.714	0.100
Qatar	4387	2390.128	27	0	0.832	0.776	0.141
Romania	5498	3677.699	27	0	0.859	0.812	0.157
Russian Federation	4882	2552.385	27	0	0.947	0.929	0.138
Saudi Arabia	4331	1961.225	27	0	0.818	0.758	0.129
Singapore	5921	4356.157	27	0	0.941	0.922	0.165
Slovenia	4389	2602.438	27	0	0.914	0.886	0.147
Sweden	5472	3812.362	27	0	0.957	0.942	0.160
Syrian Arab Republic	4370	1530.232	27	0	0.807	0.743	0.113
Thailand	6066	5928.414	27	0	0.592	0.456	0.190
Tunisia	5098	2293.086	27	0	0.864	0.819	0.128
Turkey	6903	2966.255	27	0	0.858	0.811	0.126
Ukraine	3371	1064.494	27	0	0.941	0.921	0.107
United Arab Emirates	13987	4873.403	27	0	0.888	0.851	0.113
United States	10346	6840.760	27	0	0.920	0.893	0.156

Exhibit 12: Fit statistics for the Model 4 factor analysis of Students Confident in Mathematics Scale, Eighth Grade. Model 4 allows covariance between error terms of reverse directional items.

Country	Number	X^2	Degrees of Freedom	P-Value	CFI	TLI	RMSEA
Armenia	5767	721.695	21	0	0.974	0.955	0.076
Australia	7388	2070.327	21	0	0.944	0.904	0.115
Bahrain	4579	753.388	21	0	0.951	0.917	0.087
Chile	5807	1434.544	21	0	0.959	0.930	0.108
Chinese Taipei	5037	1543.665	21	0	0.986	0.976	0.120
England	3814	1259.665	21	0	0.941	0.898	0.124
Finland	4216	1645.467	21	0	0.979	0.963	0.135
Georgia	4527	837.642	21	0	0.947	0.910	0.093
Ghana	7214	183.214	21	0	0.982	0.969	0.033
Hong Kong SAR	3981	1771.480	21	0	0.962	0.935	0.145
Hungary	5165	1080.941	21	0	0.983	0.970	0.099
Indonesia	5761	679.252	21	0	0.952	0.918	0.074
Iran, Islamic Rep. of	6017	407.445	21	0	0.984	0.972	0.055
Israel	4645	1369.890	21	0	0.941	0.898	0.118
Italy	3959	1144.490	21	0	0.980	0.966	0.116
Japan	4361	2979.167	21	0	0.930	0.881	0.180
Jordan	7588	451.691	21	0	0.976	0.959	0.052
Kazakhstan	4377	619.975	21	0	0.978	0.962	0.081
Korea, Rep. of	5163	3527.713	21	0	0.953	0.919	0.180
Lebanon	3882	271.706	21	0	0.979	0.963	0.055
Lithuania	4723	989.225	21	0	0.976	0.959	0.099
Macedonia, Rep. of	3971	487.415	21	0	0.974	0.955	0.075
Malaysia	5724	774.566	21	0	0.940	0.898	0.079
Morocco	8837	487.439	21	0	0.969	0.946	0.050
New Zealand	5191	1830.904	21	0	0.958	0.927	0.129
Norway	3827	1152.607	21	0	0.972	0.952	0.119
Oman	9430	563.318	21	0	0.964	0.938	0.052
Palestinian Nat'l Auth.	7774	390.126	21	0	0.962	0.935	0.048
Qatar	4387	566.802	21	0	0.961	0.934	0.077
Romania	5498	611.262	21	0	0.977	0.961	0.072
Russian Federation	4882	1234.061	21	0	0.974	0.956	0.109
Saudi Arabia	4331	618.691	21	0	0.944	0.904	0.081
Singapore	5921	3434.861	21	0	0.954	0.921	0.166
Slovenia	4389	1021.726	21	0	0.967	0.943	0.104
Sweden	5472	2555.478	21	0	0.971	0.950	0.149
Syrian Arab Republic	4370	200.929	21	0	0.977	0.960	0.044
Thailand	6066	523.815	21	0	0.965	0.940	0.063
Tunisia	5098	584.647	21	0	0.966	0.942	0.073
Turkey	6903	1038.324	21	0	0.951	0.916	0.084
Ukraine	3371	240.143	21	0	0.988	0.979	0.056
United Arab Emirates	13987	1753.884	21	0	0.960	0.931	0.077
United States	10346	4394.600	21	0	0.948	0.911	0.142

Exhibit 13: Latent Variable model for 'Model 5' of Students Confident in Mathematics Scale, Eighth Grade. Model 5 allows covariance between error terms of reverse directional items and error terms of "teacher items"

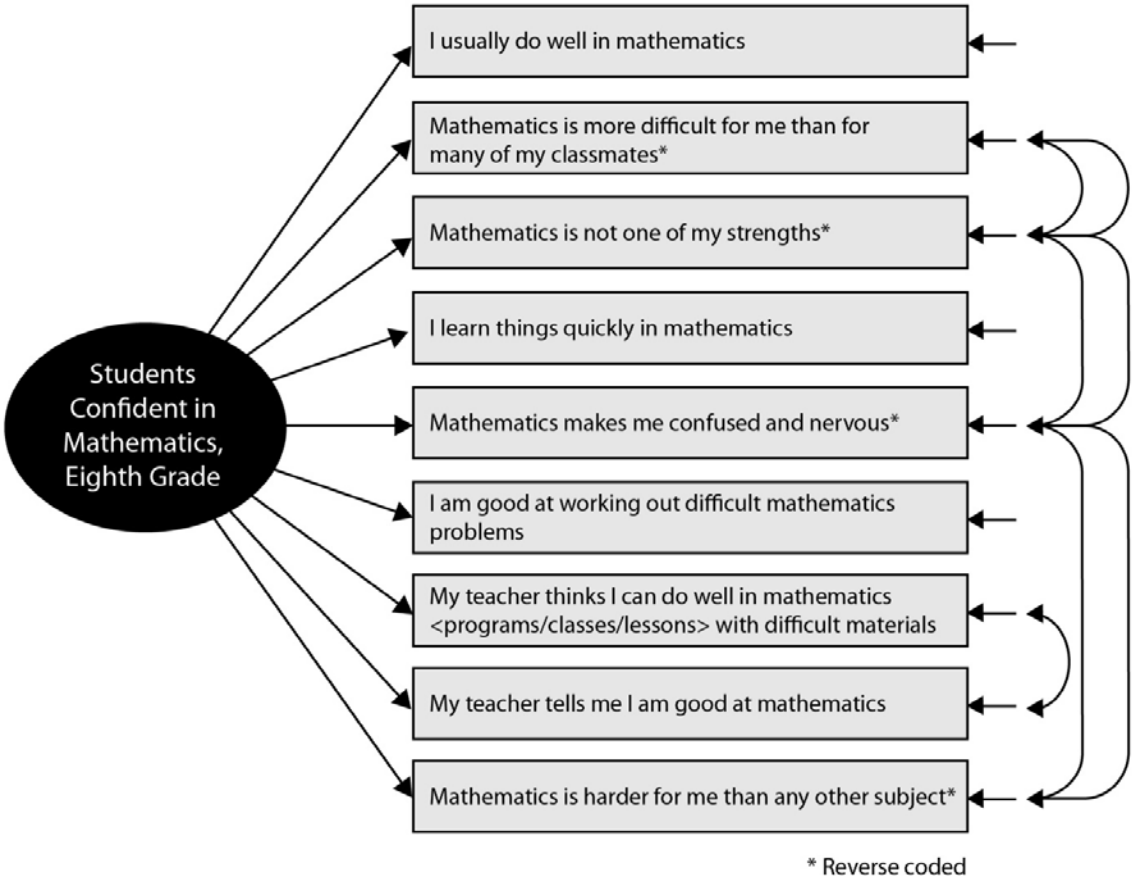


Exhibit 14: Fit statistics for the Model 5 factor analysis of Students Confident in Mathematics Scale, Eighth Grade. Model 5 allows covariance between error terms of reverse directional items and covariance between error terms of "teacher items".

Country	Number	X^2	Degrees of Freedom	P-Value	CFI	TLI	RMSEA
Armenia	5767	363.198	20	0	0.987	0.977	0.055
Australia	7388	431.283	20	0	0.989	0.980	0.053
Bahrain	4579	358.456	20	0	0.978	0.960	0.061
Chile	5807	589.616	20	0	0.984	0.970	0.070
Chinese Taipei	5037	637.915	20	0	0.994	0.990	0.078
England	3814	226.923	20	0	0.990	0.982	0.052
Finland	4216	345.753	20	0	0.996	0.992	0.062
Georgia	4527	209.378	20	0	0.988	0.978	0.046
Ghana	7214	171.173	20	0	0.983	0.970	0.032
Hong Kong SAR	3981	766.103	20	0	0.984	0.971	0.097
Hungary	5165	511.186	20	0	0.992	0.986	0.069
Indonesia	5761	488.178	20	0	0.966	0.938	0.064
Iran, Islamic Rep. of	6017	313.978	20	0	0.988	0.978	0.049
Israel	4645	281.378	20	0	0.988	0.979	0.053
Italy	3959	361.062	20	0	0.994	0.989	0.066
Japan	4361	591.464	20	0	0.987	0.976	0.081
Jordan	7588	327.382	20	0	0.983	0.970	0.045
Kazakhstan	4377	360.462	20	0	0.987	0.977	0.062
Korea, Rep. of	5163	841.558	20	0	0.989	0.980	0.089
Lebanon	3882	223.308	20	0	0.983	0.969	0.051
Lithuania	4723	462.871	20	0	0.989	0.980	0.068
Macedonia, Rep. of	3971	348.491	20	0	0.982	0.967	0.064
Malaysia	5724	369.044	20	0	0.972	0.950	0.055
Morocco	8837	455.332	20	0	0.971	0.947	0.050
New Zealand	5191	276.815	20	0	0.994	0.989	0.050
Norway	3827	332.808	20	0	0.992	0.986	0.064
Oman	9430	517.469	20	0	0.967	0.940	0.051
Palestinian Nat'l Auth.	7774	213.901	20	0	0.980	0.964	0.035
Qatar	4387	249.432	20	0	0.984	0.971	0.051
Romania	5498	284.658	20	0	0.990	0.982	0.049
Russian Federation	4882	494.236	20	0	0.990	0.982	0.070
Saudi Arabia	4331	269.755	20	0	0.977	0.958	0.054
Singapore	5921	854.860	20	0	0.989	0.980	0.084
Slovenia	4389	581.151	20	0	0.981	0.966	0.080
Sweden	5472	1098.142	20	0	0.988	0.978	0.099
Syrian Arab Republic	4370	160.788	20	0	0.982	0.968	0.040
Thailand	6066	427.962	20	0	0.972	0.949	0.058
Tunisia	5098	246.925	20	0	0.986	0.975	0.047
Turkey	6903	244.891	20	0	0.989	0.980	0.040
Ukraine	3371	77.435	20	0	0.997	0.994	0.029
United Arab Emirates	13987	618.663	20	0	0.986	0.975	0.046
United States	10346	1115.213	20	0	0.987	0.977	0.073