STUDYING MATHEMATICAL LITERACY THROUGH THE LENS OF PISA'S ASSESSMENT FRAMEWORK

Adem Ekmekci[1], Guadalupe Carmona[2]
[1]Rice University, [2]The University of Texas at San Antonio

Accurate interpretations of large-scale assessment results and sound judgments about students' mathematical literacy depend on these assessments' validity and reliability. One important type of evidence towards this validation is the dimensionality analysis, which explores the conformity between the intended factorial structure (related closely to defining a construct –e.g., mathematical literacy, and its perception) and the statistical structure of the test results. This study investigates the dimensionality of mathematical literacy in PISA. Our results show that the structural relationship between PISA's theoretical (cognitive) and score interpretation frameworks is not at an expected level. These results have important implications for the way mathematical literacy is assessed from mathematics education and psychometric perspectives.

BACKGROUND

This research focuses on the validity of assessment of mathematical literacy at a large-scale through the lens of the Programme for International Student Assessment (PISA) by studying the conformity between the intended structure of the cognitive framework provided for mathematical literacy, and the statistical structure of the results of students' scores in the 2009 implementation cycle. Based on the recommendations from the National Research Council (NRC) (NRC, 2001), the three components of assessment design: cognition, observation, and interpretation, need to be coordinated in a consistent and integrated way, as opposed to having them develop as isolated from each other. Cognition refers to the model of student learning in the domain, or mathematical literacy for our study; observation consists of the evidence provided by the student of the assessed construct; and interpretation entails making sense of this evidence. Our study is centered on the alignment between the theoretical framework for cognition and the score interpretation framework provided in PISA's 2009 assessment of mathematical literacy. There are a limited number of studies investigating the connection between the assessment framework and results. Schwab (2007) found that the multidimensional nature of PISA's science framework was reflected well in the items. Ekmekci and Carmona (2012) studied the students' responses to PISA 2003 mathematics items and detected unidimensionality for the U.S. student population. However, this study extends prior work by conducting a dimensionality analysis using the database for PISA 2009 for all students' mathematics literacy scores from 32 countries in order to better understand the complexities of assessing mathematical literacy at a large scale.

2014. In Nicol, C., Liljedahl, P., Oesterle, S., & Allan, D. (Eds.) Proceedings of the Joint Meeting of PME 38 and PME-NA 36,Vol. 2, pp. 441-448. Vancouver, Canada: PME.

Mathematical Literacy

The conversations around being mathematically literate began in the early 80's and have continued to gain importance to this day. Furthermore, the standards that have been set for literacy (being able to read and write) have sifted to incorporate mathematics as having equal importance in defining literacy (Jablonka, 2003; Moses & Cobb, 2001). In support of these views, this study is motivated by: (a) the perception of mathematical literacy through assessments; and (b) the reflection of mathematical literacy on assessments, especially in large-scale assessments whose results might have serious impact on education systems globally. In the literature, some math educators (e.g., Kilpatrick, Swafford, & Findell, 2001) focus on proficiencies or competencies when defining mathematical literacy, while others (e.g., Ojose, 2011) describe knowledge and skills. Some others (e.g., Steen, 2001) situate mathematical literacy according to its connection to real life situations (i.e., context). As diverse as multiple approaches taken by different mathematics educators and researchers might be, it seems a consensus that there are multiple dimensions or components constituting mathematical literacy. For this study, mathematical literacy is defined and viewed as "a multidimensional construct composed of distinguishable but related components rather than single, general mathematics ability" (Ekmekci, 2013, p. 1).

Since 2000, the Organisation for Economic Co-operation and Development (OECD) organizes PISA to assess 15-year olds' skills and competencies in reading, science, and mathematics through a worldwide large-scale assessment every three years. In its theoretical (cognitive) framework, PISA presents mathematical literacy as a multidimensional construct. The following is the program's given definition of mathematical literacy.

> An individual's capacity to identify, and understand, the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned, and reflective citizen. (OECD, 2003, p. 24).

PISA's mathematical literacy framework has a multidimensional structure composed of three main attributes: *content*, *processes* and *context*. Content is divided into four sub-dimensions: *quantity*, *space*, *shape*, and *change and relationship*. The process dimension has three sub-dimensions: *reproduction*, *connections*, and *reflection*. Context is composed of four sub-dimensions: *personal*, *educational/occupational*, *public*, and *scientific*. The goal of this study is to show how and to what extent this multidimensional structure is reflected on the actual tests by analyzing dimensionality of the students' responses to PISA 2009 mathematics items for 32 countries participating in the OECD.

Test Dimensionality

One of the most powerful ways to explore the connection and conformity between the framework for mathematical literacy and its assessment is dimensionality analysis. Dimensionality of a test could be informally defined as "the minimum number of examinee abilities measured by the test items" (Tate, 2002, p.182). If items in a test are found to have a unidimensional structure, then this set of items are said to be measuring one dimension of a construct. Similarly, if an assessment is said to be measuring several important attributes of a construct, then it is expected to have a multidimensional structure. Issues in development and use of large-scale assessments such as validity and test fairness are related to test dimensionality. For example, unidimensionality is one of the basic assumptions of some measurement models such as Rasch model (Hattie, 1985). The results of the tests whose items are calibrated according to these measurement models have to produce a unidimensional structure for construct validation of those tests (Rubio, Berg-Weger, & Tebb, 2001). However, it might be the case that a test that is intended to be unidimensional measures more than one latent variable (construct or one dimension of a construct). Conversely, it might be the case that some factors that do not relate to construct being measured, such as item type and format, could introduce multidimensionality to the assessment. Therefore, analyzing the dimensionality of an assessment is important and required for construct validity and to ensure accurate interpretations of test results.

PROBLEM STATEMENT

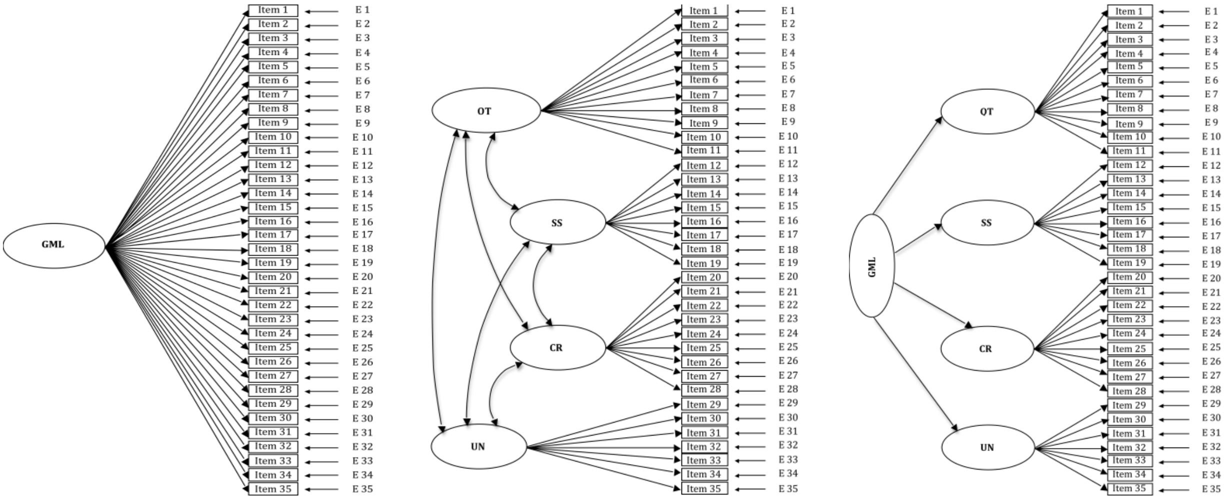
The dimensionality of PISA's mathematical literacy assessment with the inclusion of data from 32 OECD member countries has not been undertaken before. Thus, this investigation is an important contribution to the study of its construct and inferential validity. Moreover, assessing dimensionality of PISA mathematics items is needed to understand the relationship between the important assessment design components of PISA's assessment design for mathematical literacy, as recommended by the NRC (NRC, 2001). The significance of this study comes from the need to provide evidence for validation process of PISA's mathematical literacy assessment. Prior studies (e.g., Ekmekci & Carmona, 2012; Schwab 2007) have set the ground in this direction. However, this study extends prior work by conducting a comprehensive dimensionality analysis incorporating all students' responses to mathematics items from 32 OECD member countries in order to better understand the complexities of assessing mathematical literacy globally and at a large scale. The following are the research questions that guided this study.

1. What is the correspondence between the dimensional structure of PISA's mathematical literacy assessment framework and its score interpretation framework in terms of the content, process, and context dimensions?
2. What is the best representation for the dimensional structure of the PISA mathematics items used to assess students' mathematical literacy?

METHODS

This study entails a secondary-analysis of the dataset from the OECD's PISA database. The data includes student responses to individual mathematics items from 32 OECD member countries in PISA 2009. There is a variety of ways to test dimensionality of tests (see Hattie, 1985, for a comprehensive list). Having a well-developed mathematical literacy framework in PISA means that there is a strong prior expectation about the factorial structure of mathematics items (multidimensionality). In presence of a prior expectation, confirmatory factor analysis (CFA) is considered the best approach to analyze the structure of the assessed construct, i.e., mathematical literacy (Kline, 2010; Tate, 2002).

Seven CFA models were developed based on the mathematical literacy dimensions described in OECD's assessment framework for mathematical literacy. These models include a unidimensional model, three (content, process, and context) correlated factor (1-level) models, and three (content, process, and context) higher order factor (2-level) models. Correlated factors of 1-level models and factors at the first level of level-2 models are the same factors – the sub-dimensions of each main dimension. The latent factors for content dimension are thus *quantity*, *space*, *shape*, and *change and relationship*. The factors for process dimension are *reproduction*, *connections*, and *reflection*. Lastly, the factors for context dimension are *personal*, *educational/occupational*, *public*, and *scientific*. Sample illustrations for different types of models are given in Figure 1 below.



Model 1 (unidimensional Model 2 (Level-1 Content Model 5 (Level-2 Content

GML: General Mathematical Literacy, QT: Quantity, SS: Space & Shape,
CR: Change & Relationship, UN: Uncertainty, E: Error Term.

Figure 1: Sample models for the content dimension.

Each CFA model was tested with the student responses to mathematics items. There were 35 mathematics items in PISA 2009. They were dichotomously scored (correct and incorrect). The binary nature of the response data requires using a weighted least squares means and variance adjusted (WLSMV) estimator for CFA (Muthen & Muthen, 2012). The total number of respondents from 32 OECD member countries was 276,142. This large sample size could inflate the power of chi-square tests on which CFA analyses were based (Kline, 2010). Therefore, to avoid Type-I error, a smaller sample was derived randomly using appropriate sampling weights to avoid any bias in the selection. Since the number of mathematics items were large compared to typical CFA analyses, a minimum of 15,000 observations were needed (lower number of observations produced incomplete matrices for CFA calculations). This minimum number also met the criteria for minimum sample size (at least three to five times the number of correlations between items) for CFA with dichotomous items in the literature (Tate, 2002).

The statistical software Mplus 6.12 (Muthen &Muthen, 1998-2011) was used to conduct confirmatory analyses (with WLSMV being the default estimator for categorical data). For each of the three dimensions, the factorial structure of the students' responses and the assessment framework were expected to corroborate each other. This would provide evidence supporting construct validity of the PISA assessment. In other words, multidimensionality was expected in the response data. The first research question addressed how different factorial models (derived from the PISA's mathematical literacy framework) would fit the students' responses to PISA mathematics items. Goodness of fit indices (GFIs) obtained from CFA analyses such as comparative fit index (CFI), the Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA) were used to evaluate model-fit. Moreover, individual item parameter estimates (factor loadings and R-square values) were evaluated to see how each mathematics item behaved in each model (i.e., the connection between items as observed indicators and their related dimensions as latent factors).

The second research question related to comparing different structural models in order to find the best models that represented the dimensionality of response data. DIFFTEST (alternative version of chi-square difference testing modified for WLSMV estimator) and ΔGFI methods were used to compare models within each three main dimensions (content, process, and context).

Hypotheses

The single-factor model (Model 1) illustrates the hypothesis that PISA mathematics items measure a single construct labelled as general mathematical literacy (GML). The second type of models (Models 2-4) embody the hypothesis that the PISA mathematics items helps explain mathematics knowledge, competencies, and skills in terms of correlated factors of related dimension (content, process, or context) as the latent constructs. The third type of models (Models 5-7) illustrates the hypothesis that the PISA mathematics items measure GML (level-2 factor) by factors (the level-1 latent variables) of related dimension (content, process, or context).

RESULTS

All seven models were found a good fit for PISA 2009 mathematics items. Model fit indices are given in Table 1. All of GFI indices were significant according to the criteria for those indices set by Hu and Bentler (1999). In other words, the responses to the mathematics items do not contradict any of the models proposed for the dimensionality of PISA's mathematics framework. However, high correlations between latent factors in level-1 models (with the lowest correlation coefficient of 0.860) and high latent factor loadings in level-2 models (with loadings of at least 0.841) further supported the unidimensionality. Complete table of these values will be presented in the session. Relating these results to the first research question (response-framework correspondence), overall model-fit results indicate a rather weak reflection of the mathematical literacy framework in the structural representation of the PISA mathematics items. On the other hand, since model-fit indices are relatively strong for all models, multidimensionality also holds. Therefore, results for model fit indices imply that there is evidence supporting both the unidimensionality and multidimensionality of mathematics items in terms of the content, process, and context dimensions.

Secondly, all of the individual parameter estimates were found significant in each model meaning that all models provided a good account for factor loadings and that each mathematics item plays an important role in all models. A complete summary of individual item parameter estimates will be given in the presentation session. This supports that the mathematics framework is reflected in the multi-level models of dimensionality in the PISA mathematics items with respect to the three dimensions.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
χ^2 value	743.5	711.2	741.6	729.4	713.7	742.6	731.9
d.f.	560	554	557	554	556	559	556
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.000	0.000
CFI/TLI							
CFI	0.980	0.983	0.980	0.981	0.983	0.980	0.981
TLI	0.979	0.982	0.979	0.980	0.982	0.979	0.980
RMSEA							
Estimate	0.005	0.004	0.005	0.005	0.004	0.005	0.005
90% C.I.	[0.004, 0.005]	[0.003, 0.005]	[0.004, 0.006]	[0.004, 0.005]	[0.003, 0.005]	[0.004, 0.005]	[0.004, 0.005]
Prob. ≤ 0.05	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 1: Model fit indices (all statistics are significant)

Lastly, model comparison results revealed that the 2-level model performed better with the PISA 2009 mathematics items in terms of the content and the context dimensions. Therefore, a multidimensional content and context models were more plausible than the unidimensional model. However, this is not the case for the process dimension, where the unidimensional model was preferred to the multidimensional models. Complete results of the model comparisons (including statistical values) will be presented at the conference session. The summary of results is provided in Figure 2.

Content: 2-Level Model > 1-Level Model > Unidimensional Model

Process: Unidimensional Model ≥ 2-Level Model ≥ 1-Level Model

Context: 2-Level Model > 1-Level Model > Unidimensional Model

Figure 2: Model comparisons for PISA 2009.

DISCUSSION AND IMPLICATIONS

In summary, overall results reveal that although the most robust tools identified in the literature were used for analyzing PISA's 2009 mathematics literacy test dimensionality, the results are inconclusive, and in some cases, contradictory. In other words, the connection between the assessment framework and the statistical structure of mathematics items is rather weak in that the intended multidimensional nature of mathematics items is not reflected well enough in the students' responses. PISA is one of the most widely recognized and respected assessments in the world, having a well-articulated and comprehensive mathematical literacy framework and a robust psychometric design. Yet, the major components of its assessment design are not at an expected level of corroboration. This has important implications for mathematics education, measurement, and psychometrics fields.

The authors argue that psychometric methods that are most commonly being used for large-scale assessments (e.g. Rasch models) might be too limiting to provide evidence for the types of constructs the field of mathematics education is interested in and in need of assessing, especially those with multidimensional structure. An important implication for the field of mathematics education is that this area of study is in high need of new assessment designs that can bring to bear other views on mathematics literacy -beyond those addressed in PISA, and that incorporate more current psychometric models that allow for assessment of mathematical literacy in a multidimensional manner. This more consistent alignment between the nature of mathematical literacy construct and psychometric approaches allowing for multidimensionality in assessments can provide a more encompassing perspective and more valid assessments, especially those that are implemented at a large-scale and that have such high stakes decisions in educational systems all over the world.

References

Ekmekci, A. (2013). *Mathematical literacy assessment design: A dimensionality analysis of Programme for International Student Assessment (PISA) mathematics framework* (Doctoral dissertation). Retrieved from http://hdl.handle.net/2152/21338

Ekmekci, A., & Carmona, G. (2012). Mathematical literacy assessment design: A multivariate analysis of PISA 2003 mathematics items in the U.S. In L. R. Van Zoest, J. J. Lo, & J. L. Kratky. (Eds.), *Proc. 34th annual meeting of the North American Chapter of the Int. Group for the Psychology of Mathematics Education* (p. 390). Kalamazoo, MI: Western Michigan University.

Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9*(2), 139-164.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Jablonka, E. (2003). Mathematical literacy. In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 77-104). Berlin: Springer.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Moses, R. P., & Cobb, C. E. (2001). *Radical equations: Civil rights from Mississippi to the algebra project*. Boston, MA: Beacon.

Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus, version 6.12* [computer software]. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academies Press.

Ojose, B. (2011). Mathematics literacy: Are we able to put the mathematics we learn into everyday use? *Journal of Mathematics Education*, *4*(1), 89-100.

Organisation for Economic Co-operation and Development (OECD). (2003). *The PISA 2003 assessment framework – Mathematics, reading, science, and problem solving knowledge and skills*. Paris: OECD Publishing.

Rubio, D. M., Berg-Weger, M., & Tebb, S. S. (2001). Using structural equation modeling to test for multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(4), 613-626.

Schwab, C. J. (2007). *What can we learn from PISA? Investigating PISA's approach to scientific literacy* (Unpublished doctoral dissertation). University of California, Berkeley, U.S.

Steen, L. A. (2001). *Mathematics and democracy: The case for quantitative literacy*. Princeton, NJ: National Council on Education and the Disciplines.

Tate, R. (2002). Test dimensionality. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation*. Mahnah, NJ, London: L. Erlbaum.