Psychometric Examination of the Mathematics Skills Self-Efficacy Scale

University of Kentucky
College of Education

Zijia Li, MS, Michael D. Toland, PhD, & Ellen L. Usher, PhD

KENTUCKY P20 INNOVATION LAB
A PARTNERSHIP FOR NEXT GENERATION LEARNING

Problem

Research findings have consistently shown that students' self-efficacy beliefs are highly related to their achievement behaviors in the academic domain (Brown & Lent, 2006; Pajares, 1997; Schunk & Pajares, 2005). Self-efficacy refers to a task-specific judgment of one's capabilities. Individuals' level of self-efficacy therefore differs according to the task at hand (Bandura, 1986). Instruments designed to assess self-efficacy should be specific to given tasks and correspond with the performance/behavioral outcomes they are designed to predict (Bandura, 2006).

The Mathematics Skills Self-Efficacy Scale (MSSE) examined in our study was specifically designed for evaluating middle grades (Grade 4-8) students' judgments about their capabilities to successfully solve different types of mathematics problems. In this study, we examined the psychometric quality of the MSSE to assure its measurement precision, for the interpretation of data can only be as good as the quality of instruments used (Bond & Fox, 2001).

Previous Research

The literature review indicated that specific self-efficacy beliefs may be more meaningful than generalized beliefs. Domain-, subject-, task-, problem-, and skill-level mathematics self-efficacy beliefs were favored to predict different levels of learning outcomes (Bandura, 1977, 1986; Pajares, 1996; Pajares & Miller, 1995). Problem-related self-efficacy beliefs were strong predictors for students' performance on similar problems; students' actual grades were predicted by domain-specific self-efficacy scales.

The most commonly used methodologies in examining mathematics self-efficacy assessments were based on classical test theory approaches, Cronbach's alpha, confirmatory factor analyses, structure equation modeling, and path analyses. Categorical data obtained from Likert scales were treated as continuous variables. These traditional data analysis methods may decrease the accuracy of the results. Only a few studies used more advanced technology such as item response theory (IRT) to the psychometric properties of self-efficacy scale and items (Embretson & Reise, 2000; Toland & Usher, 2011).

Purpose of the Study

The purpose of our study was to evaluate the psychometric properties of the MSSE using comprehensive methods including parallel factor analysis (PFA), Rasch rating scale modeling, linear regression modeling, and Mantel-Haenszel differential item functioning (DIF). It is expected that our study will provide useful information about the MSSE, including its dimensionality, reliability, validity, and item bias information. We also examined the MSSE measurement of students' judgment of their math capabilities, variability across grade levels (i.e., 4th, 5th, 6th), and its relationship to future mathematics performance.

Method

Participants

Data were collected in fall 2010 as part of a larger longitudinal study involving students from three schools in the Southeastern United States. The sample consisted of 367 students (69 fourth, 98 fifth, and 200 sixth graders), who were 48.2% female and self-identified as White (57.8%), African American (23.4%), Hispanic (7.9%), Asian (2.7%), and other ethnicities (6.8%).

Analyses and Results

Rasch Rating Scale Model

Table 1. *Category Counts, Average Measures, Threshold (Structure, Step) Measures, and Outfit Mean-Square Statistics for a 4-Point Scale*

Category Label	Observed Count	Average Measure	Outfit Mean-Square	Thresholds
1	491	-0.38	1.24	
2	991	0.14	0.8	-0.88
3	2833	1.2	0.89	-0.34
4	5594	2.45	1.03	1.22

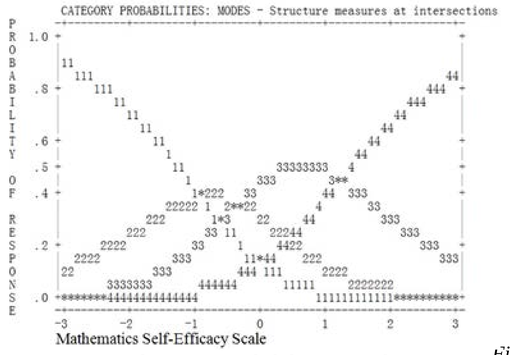

Figure 1. Final category probability curves for a 4-point scale that was initially a 6-point scale.

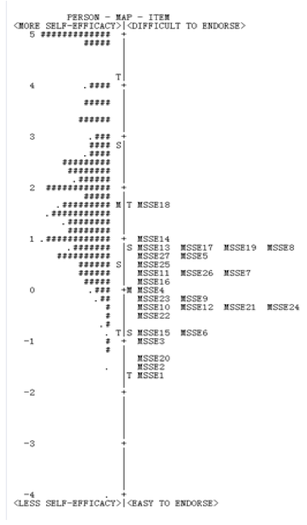

Figure 2. Variable map for a 4-point scale that was initially a 6-point scale.

Differential Item Functioning Analyses

Table 2. *Differential Item Functioning by Grade Level (4th, 5th and 6th) Using MH Approach*

	4th vs. 5th			5th vs. 6th			6th vs. 4th		
	Mantel χ^2	*p*	Class[a]	Mantel χ^2	*p*	Class[a]	Mantel χ^2	*p*	Class[a]
MSSE1	0.07	0.79	AA	1.26	0.26	AA	2.57	0.11	BB-
MSSE2	4.41	0.04	CC+	0.37	0.54	AA	4.60	0.01	CC-
MSSE3	0.49	0.48	AA	0.05	0.82	AA	0.15	0.70	AA
MSSE4	0.14	0.71	AA	5.45	0.02	CC+	7.04	0.01	CC-
MSSE5	0.24	0.62	AA	0.25	0.62	AA	0.01	0.91	AA
MSSE6	0.26	0.61	AA	4.59	0.03	CC+	5.74	0.02	CC-
MSSE7	0.01	0.93	AA	31.82	0.00	CC+	24.22	0.00	CC-
MSSE8	0.02	0.89	AA	1.96	0.16	BB+	0.83	0.36	BB-
MSSE9	0.07	0.79	AA	0.77	0.38	AA	0.22	0.64	AA
MSSE10	0.07	0.80	AA	4.80	0.03	CC-	2.85	0.09	BB+
MSSE11	0.10	0.75	AA	1.31	0.25	BB+	0.84	0.36	AA
MSSE12	0.06	0.81	AA	11.62	0.00	CC-	11.98	0.00	CC+
MSSE13	0.87	0.35	AA	1.81	0.18	BB+	0.07	0.80	AA
MSSE14	4.57	0.21	BB-	0.89	0.34	AA	5.80	0.02	CC+
MSSE15	0.50	0.48	AA	4.75	0.03	BB-	7.14	0.01	CC+
MSSE16	3.50	0.06	CC+	5.28	0.02	CC-	0.06	0.80	AA
MSSE17	1.38	0.24	BB-	2.11	0.15	BB+	0.00	0.94	AA
MSSE18	1.01	0.32	AA	4.99	0.03	BB-	8.25	0.00	CC+
MSSE19	0.65	0.42	AA	9.30	0.00	CC+	4.32	0.04	CC-
MSSE20	0.22	0.64	AA	4.05	0.04	BB-	1.92	0.17	BB+
MSSE21	0.02	0.88	BB-	1.34	0.25	BB-	1.22	0.27	BB+
MSSE22	0.85	0.36	AA	1.72	0.19	BB+	0.01	0.94	AA
MSSE23	0.04	0.83	AA	2.11	0.15	BB-	2.37	0.12	BB+
MSSE24	0.01	0.92	AA	0.85	0.36	AA	0.84	0.36	BB+
MSSE25	0.28	0.60	AA	0.02	0.90	AA	0.42	0.52	AA
MSSE26	0.51	0.47	AA	7.02	0.01	CC-	10.04	0.00	CC+
MSSE27	0.02	0.90	AA	0.31	0.58	AA	0.33	0.57	AA+

[a] Classes indicate whether the differential item functioning is negligible (AA), intermediate (BB), or large (CC).

Measures

Packets were administered to students in intact classrooms and consisted of demographic questions, the MSSE, and several other related measures. Students were asked to rate how confident they were at succeeding at exercises related to 27 mathematics topics without using a calculator on a scale from 1 (*not at all confident*) to 6 (*completely confident*). Items were placed with one stem per page such as "How confident are you that you can successfully solve mathematics exercises involving …" Mathematics achievement data for 2010 winter and 2011 spring were obtained from school records.

Key Findings

- PFA indicates two factors; however, the theoretical expectation for these items is that they belong to one underlying factor of mathematics skills. A closer inspection of the two factors reveals that they are artifacts of difficulty level. Factor 1 consisted of items that were more difficult to endorse; Factor 2 consisted of items that were easier to endorse. Also, according to Gorsuch (1983), when the ratio of the first to second eigenvalue (4.78) is greater than a value of three, this suggest essential unidimensionality among the items.
- The longer response scales (6-point) can be reduced to a more parsimonious 4-point scale with higher reliability (0.99). Using the 4-point scale, infit and outfit item statistics for all items were found to fall within the recommended range of 0.5 to 1.5, which also gives evidence for the embedded assumption of a unidimensional structure in the Rasch rating scale model.
- DIF analyses suggest that some items function differently for 4th, 5th, and 6th grade students. For example, MSSE item 14 (how confident are you that you can successfully solve problems involving equations with two or more variables) shows large functional differences between 4th and 6th graders; intermediate differences between 5th and 6th graders; and the difference between 5th and 6th graders is negligible.
- Mean comparisons show the total MSSE score of 4th graders are significantly lower than that of 6th graders; however, 4th and 5th graders and 5th and 6th graders showed similar scores.
- Linear regression results for 6th graders indicate that MSSE total score is a significant predictor of students' mathematics performance at both Winter 2010 and Spring 2011. For 5th graders, MSSE total scores only predict Winter 2010 MAP scores. For 4th graders, MSSE is not a significant predictor for students' mathematics performance on standardized test MAP scores.

Discussion

Our findings suggest that MSSE assessments (27 items) can be reduced to 4 categories with some evidence of stability and generalizability, which is similar to the findings of Smith et al. (2003), who used writing self-efficacy items with upper level elementary children (grades 4 and 5), and of Toland and Usher (2011) on similar MSSE assessments (24 items). It is important to note that the meaning of each category of the modified 4-response scale is not equal to the original 6-point version.

DIF analyses indicate that student in different grade levels may have different understandings of the Mathematics skills in MSSE. Higher level students have more comprehensive math understanding and stronger predictive power for the MSSE. Researchers interesting in assessing students' self-efficacy for using specific mathematics skills should use caution when using this measure with students of different grade levels, as they seem to interpret items differently.

*For more information, please contact Zijia Li: **zijia.li@uky.edu***