Reliability, Validity, and Measurement Invariance of Attitudinal and Social Norm Factors in Math

Soung Hwa Walker

Graduate School of Education

University of California – Riverside

900 University Avenue, Riverside, CA 92521

United States

E-mail: sounghwa@yahoo.com; sounghwa.walker@ucr.edu

Abstract

The purpose of the present study was to provide a deeper understanding of multidimensional aspects of attitudes and social norms in educational contexts. Specifically, the current investigation aimed to (1) examine reliability and validity of factorial structure of math attitudes (*Affective, Behavioral*, and *Cognitive* factors) and perceived math social norms (*Parent, Peer*, and *Teacher* factors) and (2) test whether measurement of all factors would be invariant across national groups. Three nationally representative sample data (USA, Hong Kong, and Singapore; $N = 15,019$) were obtained from the 2012 Programme of International Student Assessment (PISA) database. High values in Cronbach's Alpha, Composite Reliability scores, and factor loadings indicated strong internal consistency and convergent validity of all factors. Estimated correlations among latent factors were ranged from small to moderate in value (rs = .21 - .51, ps < .001) and each square root of average variance extracted turned out to be greater than all bivariate correlations (.71 - .85), which supported clear discriminant validity. Furthermore, the results of Multigroup Confirmatory Factor Analysis (MGCFA) yielded that measurement of the six factors was equivalent across national groups at the level of strong (scalar) factorial invariance. Implications for math education and future research are discussed.

Keywords: Attitudes, Social Norms, Validity, Reliability, Measurement Invariance

1. Introduction

Psychological traits (e.g., attitude, anxiety, motivation, personality, etc.) convey intangible characteristics that are unable to measure directly. Thus, social, educational, and behavioral scientists typically create instruments such as survey questionnaires or test items to measure those proposed theoretical concepts. The primary concerns regarding applicability of the psychological constructs in various contexts include (1) how well the instruments measure the constructs that are designed to measure and (2) whether measurement of the instruments is invariant or equivalent across different population groups or times. For example, if one wants to examine the effects of math attitudes on math outcomes across multiple groups, it is a necessary step to make sure that the instruments of attitudinal factors are measured the same way across groups prior to investigate the effects of math attitudes on math outcomes among different population samples. If measurement were not invariant across groups, conclusions of a study and/or interpretations of a research finding would be biased, weak, or misleading (Horn & McArdle, 1992; Schmitt & Kuljanin, 2008; Widaman & Reise, 1997; Yap et al., 2014).

A construct is multidimensional when it refers to several distinct but related dimensions that are treated as a single theoretical concept (Law, Wong, & Mobley, 1998). For example, a significant body of research has demonstrated that there are four distinctive aspects in self-concept construct - *academic, social, emotional*, and *physical* factors (see examples, Marsh, Ellis, Parada, Richards, & Heubeck, 2005; Marsh, & Shavelson, 1985; Shavelson, Hubner, & Stanton, 1976). In addition, it has been well documented that personality traits are multidimensional with five distinctively independent theoretical concepts, also known as 'Big Five' - *openness, conscientiousness, extraversion, agreeableness*, and *neuroticism* (see examples, Barrick & Mount, 1991; Goldberg, 1990; John & Srivastava, 1999; McCrae & Costa, 1987). Meanwhile, other psychological or hypothetical concepts such as attitudes and social/subjective norms have been treated as both unidimensional and multidimensional factors. For the purpose of the present study, social norms and subjective norms are used interchangeably.

2. Literature Review

2.1. Unidimensional Attitudes and Subjective Norms

In social and behavioral science, attitudes and subjective norms have been typically viewed as unidimensional constructs. In an attempt to understand people's intentions and behaviors, Fishbein and Ajzen (1975) postulated a theory of reasoned action (TRA). In their TRA model, *Attitudes* and *Subjective Norms* (each as a single independent variable – unidimensional factor) predict *Behavioral Intentions*; and that *Behavioral Intentions* predict behavioral outcomes (Bentler & Spekart, 1979; de Vries, Dijkstra, & Kuhlman, 1988; Fishbein & Ajzen, 1975).

In educational contexts, attitudes have been frequently treated as a single factor by being conceptualized as a learned predisposition or tendency to respond to a certain object, situation, or task either favorably or unfavorably (Aiken, 1970; Fishbein, 1967). For example, using 107 independent studies from 1966 to 1993, Ma and Kishor (1997) conducted a meta-analysis to examine the relations between attitudes toward math (i.e., as a unidimensional-independent variable) and achievement in math (i.e., dependent or outcome variable). Findings showed an overall weighted mean effect size ($r = .12$) that was positive but not strong; however, the findings from separate analyses by gender ($r = .26$ for males and $r = .23$ for females) are rather substantial. Although the magnitudes of these effect sizes may be viewed as relatively

small in terms of statistical value, the practical implication is still meaningful and important (Cohen, 1988; Ma & Kishor, 1997; Rosenthal & Rubin, 1982). That is because the effect size of $r = .23$, according to the Binomial Effect Size Display (BESD), is equivalent to increasing math achievement rate from 38.5% to 61.5% (Rosenthal & Rubin, 1982). In other words, research has suggested that when individuals hold favorable attitudes toward certain tasks, they are more likely to engage in those targeted tasks and tend to put more effort into completing their work with high quality, which contributes to positive behavioral outcomes.

More recently, Lipnevich, MacCann, Krumm, Burrus, and Roberts (2011) utilized Ajzen's (1991) Theory of Planned Behavior (TPB) model to examine the relations between math attitudes and math achievement for 8th graders with the participants from USA ($N = 382$) and Belarus ($N = 339$). The TPB model represents an extended version of the TRA model with an added control factor (Ajzen, 1991). As I previously stated, in the TRA model, *Attitudes* and *Subjective Norms* predict *Intention*, and that *Intention* predicts behavioral outcomes. According to the TPB model, *Attitudes, Subjective Norms,* and *Perceived Control* predict *Intentions*; *Intentions* and *Perceived Control* jointly predict behavioral outcomes. Although the authors did not investigate the direct association between *Attitudes* and behavioral outcomes (i.e., math achievement), they found that, among the TPB model components, the *Attitudes* factor was the strongest predictor of Math *Intentions* for both USA and Belarusian 8th graders (Lipnevich et al., 2011).

2.2. Multidimensional Attitudes and Subjective Norms

Regarding math education in particular, Neale (1969) explained math attitudes from the multidimensional perspective with four categories, "a *liking or disliking* of mathematics, a *tendency* to engage in or avoid mathematical activity, a *belief* that one is good or bad at mathematics, and mathematics is *useful or useless*" (p. 632). McLeod (1992) asserted that there might be three components in attitudes; *Affect, Emotion*, and *Belief.* Some of the most widely known math attitudes measurement research includes Tapia's (1996) Attitudes Toward Math Inventory (ATMI) which identified and measured four elements of math attitudes; *Enjoyment*, *Value*, *Self-Confidence*, and *Motivation* (Lim & Chapman, 2013; Majeed, Darmawan, & Lynch, 2013; Tapia & Marsh, 2002). Meanwhile, Zan and Di Martino (2014) posited a three-factor model of math attitudes; i.e., "*Affective, Cognitive*, and *Behavioral*" factors (p. 2). More recently, Khine, Mutawah, and Afari (2015) conceptualized that the three sub-domain attitudinal factors (i.e., *Liking*, *Value*, and *Confidence*), independently, would predict math achievement. Upon surveying 387 high school students from the United Arab Emirates, Khine et al. (2015) found that 36% of the variance in math achievement ($R^2 = .36$) was explained by the *Confidence* factor alone (p. 207).

Subjective norms present the perceived social pressure to engage or perform certain tasks or activities from people whom s/he thinks highly of, cares for, or respects the most (Ajzen, 1991; Bentler & Speckart, 1979). As multidimensional approaches to subjective norms in educational contexts, Wentzel (1998) examined the relations between perceived social supports and various motivational outcomes with young American adolescents ($N = 167$). The results of her study finding showed that (1) perceived peer support (*Peer* factor) turned out to be the strongest predictor of prosocial goal pursuit ($\beta = .29$, p < .001); (2) perceived teacher support (*Teacher* factor) was the strongest predictor of school interest ($\beta = .33$, *p* < .001); and, (3) family cohesion (*Family* factor) predicted mastery orientation the most ($\beta = .23$, *p* < .001) among motivation variables (Wentzel, 1998, p. 206). On a global scale, Walker (2017) investigated how each subjective norm factor (i.e., *Parent, Friend*, and *Teacher*), separately, would be associated with academic intentional behaviors in math across

national groups. The *Teacher* factor turned out to be the strongest predictor of math academic intention among the three subjective norm factors (*ßs* = .14, *p* < .001) for USA, Germany, Japan, and Korea. Developmentally speaking, children and young adolescents tend to pursue more independence socially, psychologically, and emotionally. Thus, healthy social supports from various forms of subjective norms (e.g., parents, peers, and teachers) during the crucial transition period from young adolescence to adulthood might greatly impact students' perceptions of academic motivation and achievement and these students' long-term views on education (Phelan, Davidson, & Cao, 1992; Wentzel, 1998).

2.3. Present Study

To update methodological and theoretical development of math attitudes and social norms in a global context, the present study aimed (1) to evaluate factorial structure of a multidimensional representation of math attitudes and perceived math social norms and (2) to test if the factors display measurement invariance across national groups (USA, Hong Kong, and Singapore) using internationally representative data from the 2012 Programme of International Student Assessment (PISA). The findings of the present study may add new insights into educational and behavioral-social science research on attitudes and social norms in terms of theoretical, methodological, and practical applicability. Figure 1 shows a visual representation of the proposed conceptual model to be tested. Math attitudes are comprised with the three distinctively separate factors; (a) the *Affective* factor that describes emotional responses to math, (b) the *Behavioral* factor that describes academic behavioral tendencies related to math, and (c) the *Cognitive* factor that describes one's belief about his/her capabilities in math. Multidimensional perceived math social norms also consist of the three independent factors; (a) *Parent,* (b) *Peer*, and (c) *Teacher.*
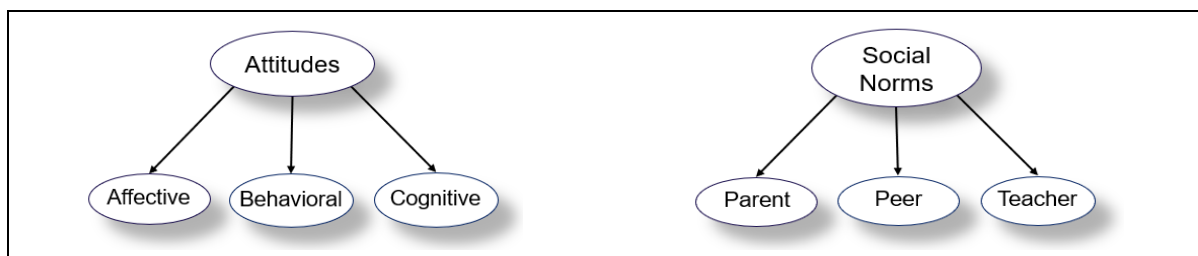

Figure 1. Conceptual models for the present study.

3. Methodology

3.1. Data Source

The dataset to be used for the present study was drawn from the Student Background Questionnaires (SBQ) of the 2012 PISA. Among 65 participating OECD countries around the world in 2012, three countries (USA, *N* = 4978; Hong Kong, *N* = 4670; and Singapore, *N* = 5546) were selected for the current investigation of factorial structure and measurement invariance. The selected 2012 PISA SBQ items for the present study (math attitudes and math social norms) were reported on 4-point Likert-type scales (i.e., 1 = strongly agree, 2 = agree, 3 = disagree, and 4 = strongly disagree). However, for the purpose of simplicity and convenience, all measured items were rescaled (0 = strongly disagree, 1 = disagree, 2 = agree, and 3 = strongly agree) to show that the higher number would represent the greater numerical value of the students' self-reported survey responses. In other words, from 0 to 3, values are close to 3 indicate higher mean scores. Table 1 presents descriptive statistics from all three

countries. The skewness and the kurtosis of all 20 observed variables were lower than the absolute value of 3, which implies the data were fairly normally distributed.

Table 1. Factors and Observed Items' Contents with Descriptive Statistics

Factors	Items	Contents	Mean	SD	Skew	Kurt
Math Attitudes						
Affective	af1	I enjoy reading about math	1.48	.87	-.02	-.67
Factor	af2	I look forward to my math lessons	1.65	.83	-.16	-.68
	af3	I do math because I enjoy it	1.62	.94	-.11	-.89
	af4	I am Interested in the things I learn in math	1.71	.87	-.20	-.66
Behavioral	bh1	I finish my math homework in time	2.01	.77	-.46	-.13
Factor	bh2	I work hard on my math homework	2.00	.73	-.40	-.01
	bh3	I study hard for math quizzes	1.66	.79	-.06	-.47
Cognitive	cg1	If putting enough effort, I can succeed in math	2.46	.62	-.89	-.67
Factor	cg2	I do well in math is completely up to me	2.25	.74	-.70	-.06
	cg3	If I wanted to, I could do well in math	2.33	.69	-.82	-.50
Perceived Math Social Norms						
Parent	pt1	My parents believe math is important to study	2.31	.65	-.64	.45
Factor	pt2	My parents believe math is important for career	2.17	.72	-.55	.05
	pt3	My parents like math	1.62	.78	-.10	-.40
Peer	pr1	Most of my friends do well in math	1.87	.67	-.22	.05
Factor	pr2	Most of my friends work hard at math	1.89	.67	-.30	.22
	pr3	My friends enjoy taking math tests	1.18	.78	.40	-.10
Teacher Factor	tc1	Teacher shows an interest in student's math learning	1.97	.84	-.32	-.74
	tc2	Teacher gives extra math help when students need it	2.25	.81	-.75	-.31
	tc3	Teacher helps students with their learning in math	2.35	.76	-.91	.02
	tc4	Teacher continues teaching until students understand	2.08	.90	-.59	-.64

Note. SD = Standard Deviation; Skew = skewness; Kurt = Kurtosis.

3.2. Data Analyses

Correlation analyses and confirmatory factor analysis were conducted to report reliability (i.e., internal consistency), factor loadings, and convergent-divergent validity. To test measurement invariance across groups, multigroup confirmatory factor analyses (MGCFA; Joreskog, 1971) were employed using *Mplus* 7 (Muthén & Muthén, 1998-2012) within the Structural Equation Modeling (SEM) framework. Among other techniques, MGCFA is considered to be the most powerful approach (Steenkamp & Baumgartner, 1998) and the most widely used method (Cheung & Rensvold, 2002; Meredith, 1993; Steinmetz, Schmidt, Tina-Booh, Wieczorek, & Schwartz, 2009; Widaman & Riese, 1997; Windle, Iwawaki, & Lerner, 1988) to assess measurement invariance across cultural and national groups. From an intensive review on measurement invariance research that was published between 2000 and 2007, Schmitt and Kuljanin (2008) reported that 75 out of 88 studies conducted empirical analyses of measurement invariance using Confirmatory Factor Analyses (CFA) methods. The principle of multigroup analysis in CFA as opposed to a single-group analysis is to

match (i.e., fit) factor models in several groups simultaneously (Joreskog, 1971; Lubke, Dolan, Kelderman, & Mellenbergh, 2003).

Utilizing the SEM framework, the goal of the present study aimed to test an *a-priori* defined model (see Figure 1; three-factor models for attitudes and social norms in math). To determine whether or not the proposed explicit theoretical model would be closely aligned with the data (e.g., three countries from the 2012 PISA), insignificant results of Chi-Square (χ^2) are desirable. However, Chi-Square (χ^2) testing is sensitive to large samples, which tends to reject models, even if they are merely inconsistent with the data (Marcoulides, Heck, & Papanastasiou, 2005; Raykov & Marcoulides, 2006). Thus, additional methods were used to evaluate the best fitting model such as the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993) with confidence interval and the comparative fit index (CFI; Bentler, 1990). Values close to or greater than .95 for CFI and below .06 for RMSEA suggest a good fit (Hu & Bentler 1999; Kline, 2011; Raykov & Marcoulides, 2006). Missing data were handled by using the full information maximum likelihood estimation (FIML) through *Mplus* 7. Using FIML produces less biased and more reliable results compared with listwise or pairwise deletion to deal with missing data (Widaman, 2006).

4. Results

4.1. Reliability and Convergent-Discriminant Validity

First, confirmatory factor analysis (CFA) was conducted to assess the reliability of each observed item; to check whether the selected survey items were good representations of each designated factor. Item reliability is evidence to show the proportion of item indicator variance that was explained by the latent factor. A factor loading of "less than .30 indicates that the items had weak validity" (Abu-Hilal, Abdelfattah, Alshumrani, Abduljabbar, & Marsh, 2013). According to the results of CFAs in Figure 2, all factor loadings were high (ranged from .52 to .90) and statistically significant. The fit indexes also indicated that the model was a great fit; χ^2 (158) = 3389.69, p < .001, RMSEA = .037 (90% CI, .036 - .038), CFI = .964. This suggests that the reliability of each item was good and each measured item was reasonably well corresponded to the specified latent construct, which also supports the convergent validity (i.e., multiple indicators of the same factor are in good alignment).
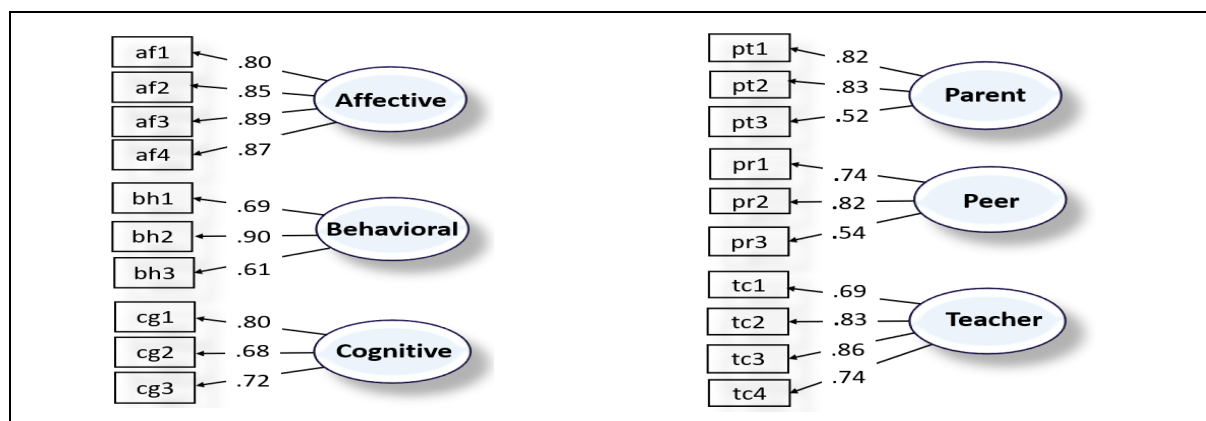

Figure 2. All factor loadings are standardized coefficients, *p*s < .001

Next, correlation analyses for all variables were conducted to check validity of the six factors. In table 2, the bolded values in highlighted are to show strong inter-correlations within factors; correlation value of affective factor items (e.g., af1 with af2, *r* = .67) shows bigger

than that of affective and behavioral factor items (e.g., af1 with bh1, $r = .21$). That is, correlation matrix clearly displayed convergent and discriminant validity of all six factors.

Table 2. Estimated Correlations among All Variables in the Model

	1	2	3	4	5	6	7	8	9	10
1. af1	--									
2. af2	.67	--								
3. af3	.72	.75	--							
4. af4	.71	.74	.77	--						
5. bh1	.21	.27	.25	.25	--					
6. bh2	.32	.41	.38	.38	.63	--				
7. bh3	.32	.36	.34	.34	.40	.54	--			
8. cg1	.33	.37	.36	.39	.23	.29	.22	--		
9. cg2	.25	.28	.28	.29	.15	.19	.15	.54	--	
10. cg3	.30	.30	.32	.33	.16	.20	.14	.58	.51	--
11. pt1	.18	.23	.18	.24	.14	.20	.16	.30	.22	.24
12. pt2	.22	.26	.21	.29	.13	.20	.18	.28	.22	.23
13. pt3	.30	.31	.30	.33	.14	.19	.21	.21	.17	.17
14. pr1	.17	.20	.16	.18	.12	.16	.14	.16	.13	.13
15. pr2	.22	.26	.23	.24	.14	.24	.22	.19	.30	.14
16. pr3	.32	.35	.33	.32	.09	.18	.23	.11	.11	.10
17. tc1	.19	.28	.21	.24	.20	.23	.19	.18	.14	.13
18. tc2	.18	.26	.20	.22	.19	.23	.16	.20	.14	.14
19. tc3	.17	.26	.20	.21	.18	.22	.16	.20	.14	.13
20. tc4	.23	.31	.24	.26	.17	.21	.18	.20	.16	.15

	11	12	13	14	15	16	17	18	19	20
11. pt1	--									
12. pt2	.72	--								
13. pt3	.38	.40	--							
14. pr1	.20	.20	.20	--						
15. pr2	.23	.21	.20	.55	--					
16. pr3	.12	.16	.27	.42	.44	--				
17. tc1	.13	.12	.15	.15	.18	.15	--			
18. tc2	.15	.13	.12	.15	.17	.12	.57	--		
19. tc3	.16	.14	.13	.14	.17	.11	.58	.72	--	
20. tc4	.12	.13	.14	.15	.18	.14	.54	.59	.64	--

Note. all correlations are $< .001$; Bolded values are inter-correlations of items for each factor; af = Affective; bh = Behavioral; cg = Cognitive; pt = Parent; pr = Peer; tc = Teacher.

Table 3 presents factor means, averaged item intercorrelations, alpha coefficients, averaged variance extracted (AVE), and composite reliability scores to demonstrate reliability of the observed variables (i.e., item indicators). Reliability refers to the overall consistency, repeatability, or dependability of item measurement. Alpha coefficients are considered the inter-relatedness of the total set of the observed items for each latent factor. In other words, how closely related a set of the observed items would be as a factor. Thus, the values are greater than .70 suggests good evidence of reliability (Cronbach, 1951). Although Cronbach's alpha alone has been widely used to evaluate internal consistency, alternative method,

Fornell-Larker's (1981) criterion, was also employed to demonstrate the level of validity and reliability of the six factors. From the classical test theory approaches to reliability of measurement model (i.e., relations between observed and latent variables), composite reliability scores were calculated - the ratio of squared sum of factor loading scores to squared sum of factor loading scores plus sum of error scores, $(\Sigma\lambda)^2/[(\Sigma\lambda)^2 + (\Sigma\varepsilon)]$ (Fornell & Larker, 1981, p. 45). The results of both Cronbach's alpha scores (.73 - .91) and composite reliability scores (.75 - .91) support the strong reliability and convergent validity of the factorial structure of multidimensional math attitudes and perceived math social norms scales.

Table 3. Factor Mean, Averaged Item Correlation, and Reliability Scores

Factors	Mean (S.E.)	mean r	Cronbach's α	AVE	CR
1. Affective (4 items)	1.48 (.009)	.73	.91	.73	.91
2. Behavior (3 items)	2.01 (.008)	.52	.77	.55	.78
3. Cognitive (3 items)	2.46 (.006)	.54	.78	.54	.78
4. Parent (3 items)	2.31 (.006)	.50	.75	.54	.77
5. Peer (3 items)	1.87 (.007)	.47	.73	.50	.75
6. Teacher (4 items)	1.97 (.008)	.61	.86	.61	.86

Note. S.E. = standard errors; mean r = average item intercorrelation; AVE = averaged variance explained; CR = composite reliability; AVE formula = $(\Sigma\lambda^2/\Sigma\lambda^2 + \Sigma\varepsilon)$; CR formula = $(\Sigma\lambda)^2/[(\Sigma\lambda)^2 + (\Sigma\varepsilon)]$; Alpha formula = $k\bar{r}/[1 + (k-1)\bar{r}]$, where k is the number of items and $\bar{r}$ is the average correlations of the items. Classical test theory approaches to reliability is the ratio of true score variance to observed score variance; observed score variance = true score variance + error score variance).

Table 4 presents estimated correlations and the values of the square roots of AVEs for the six latent variables. Correlations among six factors displayed relatively small to moderate in value (r = .21 - .51, ps < .001), which indicates no concerns of multicollinearity. This also supports discriminant validity. Discriminant validity aims to test whether concepts (i.e., latent factors) that should not be related, in fact, are not related after all (see also Table 2 for the entire 20-item correlation matrix). In addition, each square root of AVE score (.71 - .85) is greater than any bivariate correlations of latent factors (.21 - .51), which further confirms that discriminant validity requirements are satisfied (Fornell & Larker, 1981, p. 46). That is, the factorial structure clearly suggests that there are three distinctively independent sub-components in math attitudes (i.e., *Affective, Behavioral*, and *Cognitive)* and perceived math social norms (i.e., *Parent, Peer*, and *Teacher)*.

Table 4. Estimated Correlations and Average Variance Extracted among the Six Factors

	1	2	3	4	5	6
Math Attitudes						
1. Affective	**(.85)**					
2. Behavior	.49	**(.74)**				
3. Cognitive	.51	.36	**(.74)**			
Perceived Math Social Norms						
4. Parent	.33	.27	.40	**(.74)**		
5. Peer	.38	.32	.27	.34	**(.71)**	
6. Teacher	.33	.32	.28	.21	.28	**(.78)**

Note. Bolded values in diagonal with parentheses are square root of AVEs

4.2. Measurement Invariance

Within the SEM framework using *Mplus* 7, multigroup confirmatory factor analyses (MGCFA; Joreskog, 1971) were conducted to test measurement invariance across national groups. In order to avoid the possibility of getting the results by chance, each country's sample was divided into two sets of sub-data. The first half random samples (i.e., derivation) were used for the initial analyses and the other half samples were employed to cross-validate the results of the derivation samples. Table 5 presents the results of a sequence of nested models (configural, weak/metric, and strong/scalar) for both derivation and cross-validation samples of the three countries from 2012 PISA data (i.e., USA, Hongkong, and Singapore).

Table 5. Summary of a Sequential Measurement Invariance Testing

Model	χ^2	df	RMSEA (90% CI)	CFI	TLI
First Half Samples (Derivation)					
Model A	1119.24	462	.036 (.034 - .037)	.967	.959
Model B	2114.87	490	.036 (.035 - .038)	.963	.957
Model C	2778.36	517	.042 (.040 - .043)	.948	.943
Second Half Samples (Cross-Validation)					
Model A	1849.21	462	.035 (.033 - .036)	.966	.959
Model B	2010.74	490	.035 (.034 - .037)	.963	.957
Model C	2701.81	517	.041 (.040 - .043)	.947	.942

Note. Model A = Configural invariance (Same pattern of fixed and free loadings across groups); Model B = Weak/Metric invariance (Same pattern of factor loadings across groups, same constructs implied in all groups); Model C = Strong/Scalar invariance (Same item indicators/intercepts with same factor loadings across groups, allows to evaluate (compare) latent factor mean differences across groups); df = degree of freedom; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; CFI = Comparative Fit Index; TLI = Tucker Lewis Index or Non-normed Fit Index (NNFI)

4.2.1. Configural Invariance

In the first step (Model A), no equality constraints were imposed to test configural invariance across groups. Model A was used as a baseline for model comparisons to Model B and C. The initial model that assessed configural invariance (Model A) resulted in an excellent fit to data, RMSEA = .036; CFI = .967 for the derivation and RMSEA = .036; CFI = .966 for cross-validation, respectively. This means that the measurement of the factorial structure for all six factors (*Affective, Behavioral, Cognitive, Parent, Peer*, and *Teacher*) was equivalent or invariant across groups. In other words, the patterns of observed variable-latent factor relations were the same across groups but the factor loadings were allowed to be estimated freely.

4.2.2. Weak or Metric Invariance

In the second step (Model B), factor loadings were constrained to be equal across all three countries. The results of Model B also displayed excellent model fit to data. Although the fit differences in Chi-Square between Model A and B was statistically significant, $\Delta\chi^2(28) = 995.63$, *p*< .001, all practical fit indices were virtually the same (see Table 5), supporting

acceptance of the more restricted weak factorial invariance (over Model A). This means that all six factors (*Affective, Behavioral, Cognitive, Parent, Peer*, and *Teacher*) carried reasonably the same meaning to 15-year old students in all three countries. Obtaining *Metric* or *Weak* Invariance suggests that group comparisons of factor variances and covariances are acceptable, but not for group comparisons of factor means.

4.2.3. Strong or Scalar Invariance

In the third step (Model C), invoking factor loadings to be equal across groups in Model B, additional equality constraints in item intercepts were imposed across groups (i.e., constraining all intercepts to be equal in three countries). Non-invariance in item intercepts might be indicative of potential measurement bias and suggests that unknown circumstances might have influenced the way participants responded to certain survey items either too high or too low. On the other hand, establishing *Strong* invariance implies that participants, who had the same meaning/view on the latent factors (i.e., *Affective, Behavioral, Cognitive, Parent, Peer*, and *Teacher*), had actually responded to the items in the same way or very similarly across national groups.

The results of Model C again yielded a reasonably good fit to data; RMSEAs were less than .06 (derivation = .042 and cross-validation = .041) and CFIs were close to .95 (derivation = .948 and cross-validation = .947). This supports acceptance of the more restricted *Strong* factorial invariance over Model A and B. That is, the overall fit of the model (i.e., *Strong* factorial invariance) indicates that the six factors were measured equivalently in all three countries. Thus, any group difference in factor mean scores of *Affective, Behavioral, Cognitive, Parent, Peer*, and *Teacher* can be considered to reflect true group differences.

4.2.4. Strict Invariance

In the present study, the fourth step (strict invariance model) was not carried out. *Strict* factorial invariance would extend Model C by invoking the additional constraints on unique variances to be equal across groups. However, researchers suggest that the establishment of invariance in unique variances is not necessary because strict invariance has rarely been achieved in reality, thus, most studies addressed only configural, weak, and strong factorial invariance (e.g., Schmitt & Kuljanin, 2008; Widaman & Reise, 1997).

5. Conclusion and Future Research

The present study confirmed that the factorial structure of math attitudes and perceived math social norms are multidimensional, which is consistent with the prior research on math attitudes (Lim & Chapman, 2013; McLeod, 1992; Majeed et al., 2013; Neale, 1969; Zan, & Di Martino, 2007, 2014) and social norms (Walker, 2017; Wentzel, 1988). In addition, one of the key strengths of the present study was the use of two random samples from each country (the derivation and cross-validation samples) to test measurement invariance across groups. The results of the second half samples (i.e., cross-validation) successfully replicated the first half samples' (derivation) analyses, which suggest that the findings of the current investigation were not produced by luck or by chance. The results of MGCFA yielded that measurements were invariant across national groups at the level of configural, weak, and

strong factorial invariance, which means that all six factors (*Affective, Behavioral, and Cognitive* in math attitudes, and *Parent, Peer*, and *Teacher* in perceived math social norms) were interpreted reasonably the same way in three different countries; USA, Hong Kong, and Singapore.

The present research is important because the study findings might new insights into psychological, behavioral-social sciences, and the educational community for methodological, theoretical, and practical implications. Findings provided strong evidence of the practical conceptualization of multidimensional approaches to math attitudes and math social norms. That is, the survey items drawn from the 2012 PISA SBQs are good representations for further investigating substantive research inquiries regarding math attitudes, social norm influences, and math achievement across cultural groups. However, the present investigation used the 2012 PISA SBQs, in which samples included only 15-year old students in three different countries. Future research should consider using different age groups or different educational settings to validate or replicate the present study findings.

Acknowledgement

References

Abu-Hilal, M. M., Abdelfattah, F. A., Alshumrani, S. A., Abduljabbar, A. S., & Marsh, H. W.

(2013). Construct validity of self-concept in TIMSS's student background questionnaire: a test of separation and conflation of cognitive and affective dimensions of self-concept among Saudi eighth graders. *European Journal of Psychology of Education, 28*(4), 1201-1220. https://doi.org/10.1007/s10212-012-0162-1

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes. 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Aiken, L. R. (1970). Attitudes toward mathematics. *Review of Educational Research, 40*(4), 551-596. https://doi.org/10.3102/00346543040004551

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta‑analysis. *Personnel Psychology, 44*(1), 1-26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M., & Speckart, G. (1979). Models of attitude–behavior relations. *Psychological Review, 86*(5), 452-464. https://doi.org/10.1037/0033-295X.86.5.452

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A.

Bollen & J. S. Long (Eds.). *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. https://doi.org/10.1007/BF02310555

de Vries, H., Dijkstra, M., & Kuhlman, P. (1988). Self-efficacy: The third factor besides attitude and subjective norm as a predictor of behavioural intentions. *Health Education Research*, 3(3), 273-282. https://doi.org/10.1093/her/3.3.273

Fishbein, M. (1967). Attitude and the prediction of behavior. In M. Fishbein (Ed), *Readings in attitude theory and measurement* (pp. 477 - 492). New York: John Wiley & Sons, Inc.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *8*(1), 39-50. https://doi.org/10.2307/3151312

Goldberg, L. R. (1990). An alternative description of personality: The big-five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229. https://doi.org/10.1037/0022-3514.59.6.1216

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3), 117-144. https://doi.org/10.1080/03610739208253916

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. https://doi.org/10.1080/10705519909540118

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford Press.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426. https://doi.org/10.1007/BF02291366

Khine, M. S., Al-Mutawah, M., & Afari, E. (2015). Determinants of Affective Factors in Mathematics Achievement: Structural Equation Modeling Approach. *Journal of Studies in Education, 5*(2), 199-211. http://dx.doi.org/10.5296/jse.v5i2.7484

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3ʳᵈ Ed.). New York, NY: Guilford.

Law, K. S., Wong, C. S., & Mobley, W. M. (1998). Toward a taxonomy of multidimensional constructs. *Academy of Management Review, 23*(4), 741-755. https://doi.org/10.5465/amr.1998.1255636

Lim, S. Y., & Chapman, E. (2013). Development of a short form of the attitudes toward mathematics inventory. *Educational Studies in Mathematics, 82*(1), 145-164. http://dx.doi.org/10.1007/s10649-012-9414-x

Lipnevich, A. A., MacCann, C., Krumm, S., Burrus, J., & Roberts, R. D. (2011). Mathematics attitudes and mathematics outcomes of US and Belarusian middle school students. *Journal of Educational Psychology, 103*(1), 105-118. http://dx.doi.org/10.1037/a0021949

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within-and between-group differences and measurement invariance in the common factor model. *Intelligence, 31*(6), 543-566. https://doi.org/10.1016/S0160-2896(03)00051-5

Ma, X. & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education, 28*(1), 26-47. http://dx.doi.org/10.2307/749662

Majeed, A.A., Darmawan, G.N., & Lynch, P. (2013). A confirmatory factor analysis of the attitudes toward mathematics inventory (ATMI). *The Mathematics Educator, 15*(1), 121-135.

Marcoulides, G. A., Heck, R. H., & Papanastasiou, C. (2005). Student perceptions of school culture and achievement: Testing the invariance of a model. *International Journal of Educational Management, 19*(2), 140-152. https://doi.org/10.1108/09513540510582435

Marsh, H. W., Ellis, L. A., Parada, R. H., Richards, G., & Heubeck, B. G. (2005). A short version of the Self- Description Questionnaire II: operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychological Assessment, 17*(1), 81-102. https://doi.org/10.1037/1040-3590.17.1.81

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20*(3), 107-123. https://doi.org/10.1207/s15326985ep2003_1

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81-90. https://doi.org/10.1037/0022-3514.52.1.81

McLeod, D. (1992). Research on affect in mathematics education: a reconceptualization. In D. Grows (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 575-596). New York: McMillan Publishing Company.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-43. https://doi.org/10.1007/BF02294825

Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus User's Guide* (6th Ed.). Los Angeles, CA: Muthén & Muthén.

Neale, D. (1969). The role of attitudes in learning mathematics. *The Arithmetic Teacher*, *16*(8), 631-641.

Organization for Economic Co-operation and Development (OECD) (2014). PISA 2012 Results: Creative problem solving: Students' skills in tackling real-life problems (Volume V), PISA, OECD publishing. http://dx.doi.org/10.1787/9789264208070-en

Phelan, P., Davidson, A. L., & Cao, H. T. (1992). Speaking up: Students' perspectives on school. *Phi Delta Kappan, 73*(9), 695-704.

Raykov, T., & Marcoulides, G. (2006). *A first course in structural equation modeling* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum.

Rosenthal, R., & Rubin, D. B. (1982). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology, 9*, 395-396. https://doi.org/10.1111/j.1559-1816.1979.tb02713.x

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*(4), 210-222. https://doi.org/10.1016/j.hrmr.2008.03.003

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*(3), 407-441. https://doi.org/10.3102/00346543046003407

Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90. https://doi.org/10.1086/209528

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity, 43*(4), 599-616. https://doi.org/10.1007/s11135-007-9143-x

Tapia, M. (1996, November). *The attitudes toward mathematics instrument*. Paper presented at the annual meeting of the Mid-south Educational Research Association, Tuscaloosa, AL.

Tapia, M., & Marsh, G. E. (2002). *Confirmatory factor analysis of the attitudes toward mathematics inventory*. Chattanooga: Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Tuscaloosa, AL.

Walker, S. H. (2017). *Examining the Effects of Non-Cognitive Factors on Mathematics Achievement Across National Groups: USA, Germany, Japan, and Korea* (Doctoral dissertation, University of California, Riverside). ProQuest Dissertations Publishing (10287503).

Wentzel, K. R. (1998). Social relationships and motivation in middle school: The role of

parents, teachers, and peers. *Journal of Educational Psychology, 90*, 202-209. https://doi.org/10.1037/0022-0663.90.2.202

Widaman, K. F. (2006). Missing data: What to do with or without them. *Monographs of the Society for Research in Child Development, 71*, 42–64.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association. https://doi.org/10.1037/10222-009

Windle, M., Iwawaki, S., & Lerner, R. M. (1988). Cross-cultural comparability of temperament among Japanese and American preschool children. *International Journal of Psychology, 23*, 547-567. https://doi.org/10.1080/00207598808247785

Yap, S. C., Donnellan, M. B., Schwartz, S. J., Kim, S. Y., Castillo, L. G., Zamboanga, B. L., ... & Vazsonyi, A. T. (2014). Investigating the structure and measurement invariance of the Multigroup Ethnic Identity Measure in a multiethnic sample of college students. *Journal of Counseling Psychology, 61*(3), 437-446. https://doi.org/10.1037/a0036253

Zan, R., & Di Martino, P. (2007). Attitude toward mathematics: Overcoming the positive/negative dichotomy. *The Montana Mathematics Enthusiast, 3*, 157-168.

Zan, R., & Di Martino, P. (2014). Students' attitude in mathematics education. In *Encyclopedia of mathematics education* (pp. 572-577). Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-94-007-4978-8_146

Copyright Disclaimer