Check for updates

Original Research

SAGE Open
April-June 2025: 1–20
© The Author(s) 2025
DOI: 10.1177/21582440251339984
journals.sagepub.com/home/sgo
Sage

Validity Study on the Students' Attitudes Toward Mathematics Scale for English-Speaking Countries

Hun Won Choi[1] and Youn-Jeng Choi[1]

Abstract

This article focuses on the validity of the students' attitudes toward mathematics scale based on data from TIMSS 2019. The scale has been reported as having a three-factor structure for decades, but this study assumes the validity can appear differently depending on the country or culture. Thus, the scale should first be checked for invariance between countries. This study selected five English-speaking countries to prevent translation effects. The results of exploratory factor analysis (EFA) indicate that the factor loading values of the two items were below 0.4 in more than four countries. The EFA model with the two items deleted had a greater fit than the others in all countries using confirmatory factor analysis (CFA). A metric invariance model based on the EFA results showed the greatest fit when using multiple-group CFA. The findings suggest that two items may need revision, and the results must be interpreted considering differences in students' cultural backgrounds, even though the questionnaires were administered in the same language, which warrants further research.

Plain Language Summary

Validity study on the students' attitudes scale

In international comparative studies, responses between countries may vary due to nuances in culture or language, even when the same survey items are used. Recognizing this, this study was conducted to investigate potential validity issues in students' attitudinal scales in TIMSS, which has been implemented for over 20 years. To minimize environmental differences, only five English-speaking countries were included in the study. The results revealed low validity for two items across all countries. After removing these problematic items and reanalyzing the data, the validity of the scales significantly improved. Therefore, a review of these two items in the attitudinal scales in TIMSS is warranted to enhance.

Keywords

factor analysis, mathematics attitude scale, measurement invariance, TIMSS, validity study

Introduction

The Trends in International Mathematics and Science Study (TIMSS) has been widely used to compare mathematics and science achievement among countries. Researchers can use TIMSS to investigate the relationships among academic achievement in mathematics and science, student background, school effects, and teachers in these subjects. About 40 countries have participated in the survey every cycle since 1995 and have established educational policies based on the survey results. TIMSS conducts surveys using non-cognitive data (context questionnaires on the background information of students, teachers, school administrators, and national curricula) as well as cognitive data to measure the mathematics and science abilities of fourth and eighth-grade students. Many researchers have focused on factors affecting achievements in mathematics and science and have used

[1]Ewha Womans University, Seoul, South Korea

Corresponding Author:
Youn-Jeng Choi, Department of Education, Ewha Womans University, Education A Building Room 532, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, South Korea.
Email: younjengchoi@ewha.ac.kr

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

students' background information as key predictors. The scale that measure attitudes toward mathematics and science is popular as variable to analyze the effect of student backgrounds. Students' motivation for learning or positive emotions toward subjects (e.g., interest, confidence, endurance, etc.) enables them to learn steadily and encourage students to seek challenging goals (Greensfeld & Deutsch, 2022; Pekrun et al., 2002). By the same reasoning, if students have positive attitudes toward mathematics or science, they will be able to learn those subjects more easily than if they have negative feelings. Thus, attention should be paid to learners' motivation or emotions toward mathematics or science.

The TIMSS attitude scales toward subjects have been widely used as variables in recent decades. However, only a few researches have studied the validity of these attitude scales (Ertürk & Oyar, 2021; Eser, 2021; Liou & Lin, 2021; Liu & Meng, 2010; Oon & Fan, 2017; Reynolds et al., 2022; Sabah et al., 2013; Uyar, 2021). This study explores the internal structure of 27 items from students' attitudinal questionnaires from TIMSS 2019 and investigates their validity using factor analysis. Recent studies on the validity of the attitudinal scales in TIMSS have focused on determining whether the invariance of scales was sustained regardless of conditions such as gender, race, and country. However, the invariance between countries under the same conditions should be checked first rather than different languages or cultures. Thus, this study focuses on whether the same internal structure of factors is observed across countries after controlling for translation effects by examining measurement invariance. When measurement invariance and the validity of the attitude scale are secured among countries, the interpretation of the results of the attitude scale across countries will be more precise than before.

Theoretical Background

Trends and Alterations of the Attitude Scales

TIMSS has been used to survey the Students' Attitudes Toward Mathematics scale (the SAM scale) in 4-year cycles since 1995. The International Association for the Evaluation of Educational Achievement (IEA) has developed and updated many of the TIMSS 2019 attitudinal items, combining them into scales that measure a single underlying latent construct, as described in their technical report. Many of these scales have been continuously changed to gather useful information about learning attitudes over the last 20 years. Some items have consistently measured specific factors or constructs for seven cycles, while others have been changed or deleted while new items are inserted to reflect current research or trends (Martin et al., 2020). Appendix 1 shows the change trends and alterations of the attitudinal scales.

Overview of Research Trends of TIMSS Attitudinal Scales

Many researchers have used the SAM scale in TIMSS as a variable to explain students' achievement or motivation. Two hundred fifty-nine journal articles have used the scale as a predictor variable, and 24 journal articles have studied the validity of the attitude scale in the Web of Science as of May 2023. Among the 24 journal articles, most papers examined validity by confirming measurement invariance in specific conditions or using demographic characteristics such as gender, race, regional variation, school type, and home environment (Aditomo & Klieme, 2020; Ardic & Gelbal, 2017; M. Chen & Hastedt, 2022; Glassow et al., 2021). Eser (2021) investigated the measurement invariance of attitudinal scales based on home resources for fourth-grade students in Turkiye in TIMSS 2015. Liou and Lin (2021) also studied measurement invariance regarding attitude scales toward science using data from eighth-graders in Australia, the United States, and Taiwan from TIMSS 2019. Oon and Fan (2017) provided psychometric information on attitudinal scales such as measurement invariance, unidimensionality, optimum utilization, and item difficulty hierarchy based on the Rasch model using eighth graders from Hong Kong and Singapore in TIMSS 2011. Reynolds et al. (2022) examined the measurement invariance and cross-country comparability of attitudinal scales in international educational large-scale assessments using item response theory (IRT) modeling approach. Using data from fourth graders from 58 countries and eighth graders from 39 countries from TIMSS 2019, they concluded that the measurement invariance of item 19G was unsatisfactory ("My teacher tells me I am good at mathematics.").

Existing studies have also investigated factor structures across countries (Abu-Hilal et al., 2013; Ayob & Yassin, 2017). Liu and Meng (2010) examined the factor structure of attitudinal scales according to country and degree of achievement using data from eighth-graders in Japan, Hong Kong, Taiwan, and the USA from TIMSS 2003. Uyar (2021) evaluated the appropriateness, factor structure, and invariance of Turkish eighth-graders' attitudes toward mathematics from TIMSS 2015 using an exploratory structural equation model and confirmatory factor analysis (CFA). Uyar (2021) concluded that the scales satisfied the invariance conditions. Marsh et al. (2013) compared validity measures such as factor structures, method effects, gender differences, and convergent and discriminant validity between four Arab-speaking countries and four English-speaking countries using the TIMSS 2007 dataset. They highlighted methodological weaknesses in the TIMSS approach to these measures.

In contrast, some papers have studied methodological properties to measure validity. Ertürk and Oyar (2021)

examined the measurement invariance of attitudinal scales toward mathematics through different methods such as multiple group confirmatory factor analysis (MGCFA), multiple group latent class analysis (MGLCA), and a mixed Rasch model. They used data from eighth-grade students in the USA, Canada, and Turkiye from TIMSS 2015. They concluded that examining assumptions and considering the variable structure is necessary when deciding on the method in measurement invariance studies. Michaelides (2019) analyzed the factor structure of 18 item scales regarding mathematics and the effect of negative keying items using data from fourth graders from six European countries from TIMSS 2011. He stated that reverse-keyed items were responded differentially according to age or achievement level and suggested reconsidering the use of reverse-keyed items in surveys of young students to confirm the validity. Sabah et al. (2013) also discussed misfitting negatively worded items in TIMSS. They investigated how to validate a scale of eighth graders' attitudes toward science based on eight items from TIMSS 2007, using Rasch measurement perspectives across countries along the states of achievement. They concluded misfit items must be deleted to support the validity of the attitudinal scales.

Studies using TIMSS and other datasets have demonstrated the need to thoroughly check the validity when using reverse-keyed items in research (Bolt et al., 2020; Bulut, 2021; Bulut & Bulut, 2022; Kam & Meyer, 2023; Lindwall et al., 2012). Lindwall et al. (2012) emphasized that positively and negatively worded items did not show invariant across countries and may be interpreted differently depending on the student's cultural or linguistic background.

Many studies on the validity of TIMSS have considered the potential influence of students' gender, age, and cultural or linguistic background across countries, with most investigating the influence of demographic characteristics and various other measurement invariance (Abu-Hilal et al., 2013; Ayob & Yassin, 2017; Liu & Meng, 2010; Marsh et al., 2013). A research gap remains regarding whether linguistic similarity across countries can compensate for validity issues in international assessments like TIMSS that include positively and negatively worded items. This study seeks to address this gap in the literature.

A discussion on the validity of attitudinal scales in TIMSS has been conducted because it is a significant issue in large-scale international assessment. It is crucial to confirm the validity before using data to research and establish education policy. The TIMSS dataset is widely used in research to compare educational outputs across countries. Thus, it is necessary to check whether the validity of the scales is secured among countries before comparing them.

Validity Results From the TIMSS 2019 Technical Report

The IEA reported that the principal components analysis (PCA) was conducted to prove that it provides comparable measurements across countries, and the component loadings of each questionnaire item from the PCA were positive and substantial, indicating a strong correlation between each item and the scale in each country (Martin et al., 2020). Therefore, this study verified a single underlying latent construct as described in IEA's technical report using PCA.

TIMSS explains factors affecting students' motivation to learn and categorizes the SAM scale into "Students Like Learning Mathematics," "Students are Confident in Mathematics," and "Students Value Mathematics" within the report (Mullis & Martin, 2017). In this study, we did not use TIMSS's terms for learning attitudes, simplifying them to words such as "interest," "confidence," and "value recognition."

Methods

Datasets and Sample

Validity results vary when conditions such as gender, age, race, culture, and language differ. To ensure the validity of the attitudinal scale in international surveys, the gaps between these conditions must be reduced. This study selected datasets of eighth graders from TIMSS 2019 according to the results of Michaelides (2019) and Sabah et al. (2013), who found that responses to the reverse-keyed items were influenced by achievement, and this influence was stronger among young students, such as fourth graders, than among older students (Bolt et al., 2020; Bulut, 2021; Bulut & Bulut, 2022; Kam & Meyer, 2023; Lindwall et al., 2012). The questionnaire contained seven reverse-keyed items out of 27. Therefore, eighth graders were deemed to be the appropriate age group of students should be considered in the survey.

In addition, it is important to examine the internal structure of factors after controlling for translation effects to evaluate the validity of the scales used in various countries. This study chose a dataset from linguistically similar countries that had consistently participated in TIMSS at least three times recently to minimize differences in the linguistic meaning of items meaning and the gap in experiences taking the survey. TIMSS 2019 collected data on the first languages of students and which language they used to respond to the survey. This information is included in the language of the student context questionnaire from each country's dataset provided by the TIMSS website. We selected datasets from countries in which over 95% of students used questionnaires written in English and which presented no special issue according to TIMSS. We used data from five English-

speaking countries (Australia, England, Ireland, Singapore, and the United States) to investigate the factor structure and to prevent translation effects. The average mathematics achievement of students in these countries was significantly higher than that of the average eighth graders. Appendix 2 shows the average mathematics achievement of students in each country, taken from the TIMSS 2019 international results (Mullis et al., 2020). In addition, Appendix 3 shows the percentage of eighth-grade students who answered the questionnaire in English within the five English-speaking countries that we selected.

This study randomly divided the entire dataset from five countries into two samples, one for exploratory factor analysis (EFA) and the other for confirmatory factor analysis (CFA), following the approach used in other studies (Guo et al., 2022; Vaculíková et al., 2022; Willmer et al., 2019) based on Hair et al., (2006). The total number of samples, excluding missing values, was 26,693. The number of participants in Dataset A for EFA was 13,395, and Dataset B included 13,298 participants for CFA. Table 1 shows the number of samples from the five countries in each dataset.

Data Cleaning

TIMSS 2019 divided the SAM scale into three factors: interest (Item 16), confidence (Item 19), and value recognition (Item 20). Each attitude toward learning mathematics scale comprised nine items. All 27 items were rated on a four-point Likert scale (1 = *"Agree a lot"*, 2 = *"Agree a little"*, 3 = *"Disagree a little"*, 4 = *"Disagree a lot"*). Items with a positive connotation were reverse-coded to enable natural interpretations. We used listwise deletion for missing data, and no outliers were detected.

Analysis

We used principal component analysis (PCA) and principal axis factoring (PAF) in extraction methods EFA to reduce items to smaller subsets containing as much valuable information from the initial items as possible (Nunnally & Bernstein, 1994). EFA is conducted when the researcher does not know how many factors are necessary to explain the interrelationships among a set of indicators or items (Gorsuch, 1983; Pedhazur & Schmelkin, 1991; Tabachnick & Fidell, 2001). In this study, we attempted to re-explore the factor structure under the assumption that we do not know exactly how many factors there are. In PCA, we did not use rotation results when the factor rotation was not converged, or the factor matrix was not generated. We used multiple methods to determine the appropriate number of factors using a scree plot, the percentage of variance, and parallel analysis to evaluate whether the factor structure was achieved for these five English-speaking countries as intended by the IEA. The principal axis factoring (PAF) extraction method was also applied to investigate all 27 items regarding the factor structures of the three subscales for the SAM scale. A factor loading of 0.40 was used as a cut-off value (Hair et al., 2006), and the statistical software package used for the analyses was SPSS version 28.

Next, CFA was performed to find a suitable factor structure based on both the TIMSS 2019 technical report and the EFA results using SAS 9.4. CFA is used to assess the extent to which the hypothesized underlying structure of the construct under investigation fits the data (Nunnally & Bernstein, 1994; Pedhazur & Schmelkin, 1991). So we verified whether the assumed factor structure fits the data well. The estimation method was the maximum likelihood with Satorra-Bentler adjustments, as our datasets were not normally distributed (Satorra & Bentler, 1994).

The R package lavaan (based on R version 4.1.2) was used for multiple-group confirmatory factor analysis (MGCFA; Rosseel, 2012). MGCFA is an analytical method that verifies measurement invariance by checking whether the factor structure is the same across groups. Measurement invariance refers to whether an instrument, such as a questionnaire or items, is interpreted in the same way across different groups (F. F. Chen, 2008; Davidov et al., 2014; Horn & McArdle, 1992; Millsap, 2011). Once measurement invariance has been established, comparisons between scores or values are considered valid. The robust maximum likelihood (MLR) method was used as an estimation method because it has been demonstrated to be the most suitable method for non-normal data in previous simulations (Hirsch et al., 2018; Lei & Wu, 2015; Wurster, 2022). Various model-fit indices were computed, and the χ^2/df ratio was utilized

Table 1. The Number of Participants in Each Dataset.

Samples	Australia	England	Ireland	Singapore	USA	Total
Dataset A	4,054	1,424	1,856	2,417	3,644	13,395
Dataset B	4,094	1,394	1,861	2,375	3,574	13,298
Total	8,148	2,818	3,717	4,792	7,218	26,693

as a badness-of-fit index, as smaller values indicate a better fit (West et al., 2015). Constraints were controlled in the model parameters to evaluate the degree of invariance in the measurement model of the SAM scale in the MGCFA model. MGCFA was conducted to analyze metric and scalar invariance across groups with a configural invariant factor structure (Ding et al., 2022).

Results

Assumption Checks

Table 2 indicates the descriptive statistics for the 27 items of the SAM scale in Dataset A for EFA. The Shapiro-Wilk normality test is an appropriate method for small sample sizes ($n < 50$), although it can also handle large samples, while the Kolmogorov-Smirnov normality test is used for $n \geqslant 50$. For both tests, the null hypothesis states that data are taken from a normally distributed population. We performed both the Shapiro-Wilk and Kolmogorov-Smirnov tests to check the normality assumption. All datasets for the five countries had non-normality. Bartlett's test of sphericity for the 27 items was significant ($p < .001$). The Kaiser-Meyer-Olkin (KMO) test statistics ranged from 0.955 to 0.960 and the anti-image values for individual measures of sampling adequacy (MSA) ranged from 0.878 to 0.981. These statistics indicated that our data were suitable for performing factor analysis (see Table 3).

Examining the Validity sing Principal Components Analysis (PCA)

In the technical report, TIMSS 2019 reported that each of the three scales was extracted as a single underlying latent construct when using PCA with no rotation. We performed PCA for each factor separately to compare our PCA results to those in the technical report. However, Table 4 shows that confidence was extracted as a two-factor structure in all five countries with an eigenvalue of >1 and a scree plot, and 19E ("Mathematics makes me nervous") and 19G ("My teacher tells me I am good at mathematics") were double-loading items on the Confidence factor for all countries. In contrast, according to the data from Singapore, a two-factor structure with an eigenvalue of >1 was extracted from the "Confidence" items and the "Value recognition" items, and three items (20A ("I think learning mathematics will help me in my daily life"), 20E ("I would like a job that involves using mathematics"), and 20H ("My parents think that it is important that I do well in mathematics")) were double-loaded.

Examining the Validity Using Principal Axis Factoring (PAF)

Table 5 shows communalities from the PAF analysis using all 27 items, and the communalities of four items were below 0.4 for all five countries: 16C ("Mathematics is boring"), 19E ("Mathematics makes me nervous"), 19G ("My teacher tells me I am good at mathematics"), and 20H ("My parents think that it is important that I do well in mathematics") (see Table 5).

Also, the communality value of three items was below 0.4 for one country: 16B ("I wish I did not have to study mathematics"), 19C ("Mathematics is not one of my strengths"), and 20B ("I need mathematics to learn other school subjects"). The oblique factor rotation using the direct oblimin method was applied as the factor correlations ranged from 0.116 to 0.644, and the factor correlations between interest and confidence and those between interest and value recognition were high for all five countries (see Table 6).

Three factors were extracted in all five countries based on scree plots, parallel analysis, and total variance explained by factors, regardless of the TIMSS 2019 technical report. However, two items (19G ("My teacher tells me I am good at mathematics") and 20E ("I would like a job that involves using mathematics")) did not fit the scale TIMSS intended in more than four countries even though translation effects were controlled (see Table 7). Also, most of the loading values of these items were below 0.4. Moreover, three items (19F ("I am good at working out difficult mathematics problems"), 19G ("My teacher tells me I am good at mathematics"), and 20E ("I would like a job that involves using mathematics")) were assigned incorrectly to another factor, and the incorrectly assigned loading values of the three items (19F, 19G, and 20E) were above 0.40 in Singapore. Appendix 4 includes the structure factor matrix for all five countries for more information.

Examining Validity Using Confirmatory Factor Analysis (CFA)

Table 8 indicates the descriptive statistics for the 27 items of the SAM scale for Dataset B (see Table 8).

We performed four different CFA models to find the best-fitted model for our data. Because the values of average variance extracted (AVE) and composite reliability (CR) were all greater than 0.9, they were suitable for CFA (Bagozzi & Yi, 1988; Fornell & Larcker, 1981). Model A was based on the TIMSS 2019 technical report and comprised a three-factor structure using all items. Model B was composed of a single-factor structure using all items, as researchers have often used these items as one factor in their studies. Model C was based on the

Table 2. Descriptive Statistics of the SAM Scale (Dataset A).

Item no.	Item statement	Australia (*N* = 4,054)		England (*N* = 1,424)		Ireland (*N* = 1,856)		Singapore (*N* = 2,417)		USA (*N* = 3,644)	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Interest											
16A*	I enjoy learning mathematics	2.79	0.98	2.79	0.92	2.75	0.99	3.11	0.88	2.84	1.00
16B	I wish I did not have to study mathematics	2.73	1.10	2.78	1.14	2.63	1.14	2.77	1.08	2.57	1.11
16C	Mathematics is boring	2.46	1.22	2.38	1.19	2.41	1.19	2.76	1.05	2.51	1.33
16D*	I learn many interesting things in mathematics	2.80	0.92	2.73	0.88	2.70	0.96	3.05	0.82	2.85	0.96
16E*	I like mathematics	2.70	1.02	2.66	0.97	2.69	1.04	2.96	0.95	2.74	1.06
16F*	I like any schoolwork that involves numbers	2.32	0.94	2.30	0.93	2.37	0.97	2.56	0.91	2.33	0.98
16G*	I like to solve mathematics problems	2.54	1.00	2.52	0.99	2.46	1.03	2.76	0.98	2.55	1.04
16H*	I look forward to mathematics class	2.32	0.98	2.28	0.96	2.28	0.99	2.65	0.95	2.47	1.05
16I*	Mathematics is one of my favorite subjects	2.29	1.13	2.13	1.08	2.28	1.14	2.70	1.13	2.50	1.18
Confidence											
19A*	I usually do well in mathematics	3.01	0.89	3.05	0.79	3.01	0.84	2.72	1.02	3.18	0.88
19B	Mathematics is more difficult for me than for many of my classmates	2.75	1.06	2.71	1.01	2.73	1.04	2.63	1.00	2.77	1.09
19C	Mathematics is not one of my strengths	2.59	1.26	2.55	1.35	2.50	1.23	2.51	1.15	2.64	1.33
19D*	I learn things quickly in mathematics	2.73	0.93	2.76	0.91	2.71	0.93	2.71	0.94	2.86	0.96
19E	Mathematics makes me nervous	2.83	1.09	3.02	1.12	2.90	1.02	2.40	1.02	2.80	1.10
19F*	I am good at working out difficult mathematics problems	2.55	0.94	2.65	0.89	2.49	0.93	2.45	0.92	2.64	0.98
19G*	My teacher tells me I am good at mathematics	2.53	0.96	2.56	0.99	2.64	0.98	2.35	0.94	2.67	1.03
19H	Mathematics is harder for me than any other subject	2.81	1.08	2.84	1.07	2.82	1.10	2.76	1.08	2.79	1.18
19I	Mathematics makes me confused	2.50	1.03	2.51	1.07	2.41	1.03	2.43	1.01	2.52	1.12
Value recognition											
20A*	I think learning mathematics will help me in my daily life	3.29	0.84	3.12	0.90	3.03	0.99	3.16	0.85	3.14	0.92
20B*	I need mathematics to learn other school subjects	3.13	0.82	3.14	0.81	2.99	0.95	3.00	0.83	3.06	0.90
20C*	I need to do well in mathematics to get into the <university> of my choice	3.23	0.87	3.48	0.75	3.35	0.86	3.31	0.76	3.47	0.78

(continued)

Table 2. (continued)

Item no.	Item statement	Australia ($N = 4,054$)		England ($N = 1,424$)		Ireland ($N = 1,856$)		Singapore ($N = 2,417$)		USA ($N = 3,644$)	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
20D*	I need to do well in mathematics to get the job I want	3.18	0.89	3.24	0.89	3.15	0.96	3.21	0.83	3.21	0.93
20E*	I would like a job that involves using mathematics	2.43	1.01	2.30	1.03	2.27	1.04	2.47	0.99	2.40	1.09
20F*	It is important to learn about mathematics to get ahead in the world	3.23	0.82	3.12	0.87	3.04	0.94	3.21	0.79	3.22	0.87
20G*	Learning mathematics will give me more job opportunities when I am an adult	3.47	0.72	3.42	0.74	3.42	0.78	3.38	0.70	3.43	0.78
20H*	My parents think that it is important that I do well in mathematics	3.58	0.65	3.58	0.66	3.61	0.65	3.52	0.67	3.60	0.68
20I*	It is important to do well in mathematics	3.52	0.69	3.57	0.65	3.52	0.71	3.55	0.64	3.58	0.70

Note. *= reversely coded items.

EFA results; two items (19G and 20E) were deleted, and the model consisted of a three-factor structure. Lastly, Model D applied only to the data from Singapore based on EFA results because only Singapore had a different factor structure from the EFA results. Three items (19F ("I am good at working out difficult mathematics problems"), 19G ("My teacher tells me I am good at mathematics"), and 20E ("I would like a job that involves using mathematics")) were moved from each original scale to interest without deleting items.

Table 9 presents that no models were suitable for the chi-square statistics for the measurement model normalized by degrees of freedom (χ^2/df), but the values of Model C were the smallest among all models (see Table 9). Based on the three model fit indices used to find the suitable model fit (Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), and Comparative Fit Index (CFI)), Model C was the best-fitted model for our dataset from 0.050 to 0.059, which is under to the typical 0.08 cutoff according to Hu and Bentler (1999). The CFI values of Model C ranged from 0.888 to 0.917, which is close to the typical 0.900 cutoff used by Bentler (1990). Therefore, we found that Model A based on the factor structure suggested by the TIMSS 2019 technical report was not bad, but Model C, based on the factor structure of the EFA results that deleted two items (19G and 20E), fitted better than Model A. These results are meaningful in that they show the survey's effectiveness. Also, it was found that measuring the three-factor structure consisting of interest, confidence, and value recognition was desirable because of the poor fit of Model B, which was a one-factor structure. Finally, Model D only applied to Singapore's data based on the among all four CFA models. The RMSEA values of Model C ranged from 0.063 to 0.066, which is a fair fit (based on 0.05–0.08 as suggested by MacCallum et al. (1996), and the SRMR values of Model C ranged EFA results, which were moved from each original scale to the interest scale without deleting items (19F, 19G, and 20E), has a bad fit and Model C had a better fit than Model D. Thus, it was confirmed that Model C was better than Model D for Singapore.

Measurement Invariance Tests Across Five Countries Using Multiple Group Confirmatory Factor Analysis (MGCFA)

MGCFA was performed on Models A and C according to the CFA results. The model fit values of Model C were better than Model A, but both models did not have poor statistical values. Plus, Model A was proposed by TIMSS which designed the survey, and Model C was based on the EFA results. Therefore, both models were

Table 3. Bartlett's Test of Sphericity, KMO, MSA, and Cronbach's α (Dataset A).

Values	Australia	England	Ireland	Singapore	USA
N	4,054	1,424	1,856	2,417	3,644
KMO	0.960	0.948	0.955	0.957	0.958
Bartlett's test of sphericity	0.000	0.000	0.000	0.000	0.000
MSA	0.902–0.980	0.898–0.976	0.878–0.981	0.883–0.982	0.920–0.981
Cronbach'α	.940	.928	.943	.943	.936

Table 4. Number of Extracted Factors (Dataset A, PCA).

	Australia				England				Ireland				Singapore				USA			
Attitudes	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Interest	1	1	1	65.2	1	1	1	61.7	1	1	1	63.5	1	1	1	65.9	1	1	1	64.7
Confidence	1	2[a]	2	53.4	1	2[a]	2	46.1	1	2[a]	2	55.7	1	2[a]	2	55.7	1	2[a]	2	50.5
Value recognition	1	1	1	55.4	1	1	1	52.4	1	1	1	51.6	1	2[b]	1	50.7	1	1	1	54.1

Note. Extraction method: 1 = parallel analysis, 2 = eigenvalue > 1, 3 = scree plot, 4 = total variance explained by first factor (%).
[a]Double-loaded with 19E item and 19G item.
[b]Double-loaded with 20A item, 20E item, and 20H.

Table 5. Communalities (Dataset A, PAF).

Item no.	Australia	England	Ireland	Singapore	USA
16A*	0.78	0.73	0.77	0.74	0.75
16B	0.43	0.40	0.46	0.49	**0.37**
16C	**0.31**	**0.29**	**0.35**	**0.38**	**0.25**
16D*	0.54	0.52	0.51	0.52	0.53
16E*	0.81	0.78	0.82	0.81	0.82
16F*	0.64	0.57	0.57	0.56	0.64
16G*	0.71	0.69	0.62	0.73	0.79
16H*	0.69	0.65	0.62	0.63	0.71
16I*	0.72	0.67	0.72	0.78	0.75
19A*	0.58	0.52	0.56	0.62	0.55
19B	0.50	0.43	0.55	0.51	0.54
19C	0.56	**0.38**	0.47	0.72	0.47
19D*	0.60	0.50	0.56	0.52	0.57
19E	**0.26**	**0.22**	**0.27**	**0.33**	**0.32**
19F*	0.60	0.51	0.54	0.54	0.53
19G*	**0.26**	**0.18**	**0.25**	**0.33**	**0.28**
19H	0.63	0.61	0.62	0.68	0.65
19I	0.56	0.49	0.54	0.58	0.52
20A*	0.50	0.43	0.48	0.43	0.49
20B*	0.45	0.45	0.41	**0.35**	0.41
20C*	0.56	0.50	0.51	0.46	0.54
20D*	0.57	0.57	0.56	0.54	0.54
20E*	0.47	0.44	0.46	0.51	0.44
20F*	0.62	0.59	0.57	0.58	0.62
20G*	0.62	0.63	0.55	0.61	0.63
20H*	**0.32**	**0.36**	**0.26**	**0.35**	**0.37**
20I*	0.59	0.54	0.52	0.51	0.54

Note. * = reversely coded item; bold = communalities less than 0.4.

Table 6. Correlations Between Factors (Dataset A, PAF).

	Australia			England			Ireland			Singapore			USA		
Factors	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
F1	1			1			1			1			1		
F2	0.602	1		0.549	1		0.644	1		0.577	1		0.577	1	
F3	0.506	0.271	1	0.450	0.217	1	0.514	0.267	1	0.443	0.116	1	0.466	0.171	1

Note. F1 = interest; F2 = confidence; F3 = value recognition.

analyzed using MGCFA to test the measurement invariance among the five English-speaking countries with the same language conditions. These models used different datasets of observed variables (Model A consists of 27 items, and Model C consists of 25 items), so each dataset was analyzed separately and performed in the same steps. Table 10 presents the models' fit statistics and comparisons (see Table 10).

We performed a chi-square test to evaluate whether the model fits the data. However, the multiple model-fit indices must be examined because chi-square tests are substantially affected by the sample size (Bollen, 1989; Cheung & Rensvold, 2002; Meade et al., 2008; Svetina et al., 2020). Even though the chi-square test results point to incorrect specifications in all models, widely-used fit indices, namely CFI, RMSEA, and SRMR, revealed better results in all invariance conditions. As a result, RMSEA and SRMR values were less than .080. CFI yielded values ranging from 0.876 to 0.910, which is close to the typical 0.900 cutoff.

Two invariance conditions in Model C (configural and metric) were larger than the CFI of 0.90. The values of ΔRMSEA in all invariance conditions ranged from −0.002 to 0.001, those of ΔSRMR ranged from −0.004 to −0.002, and those of ΔCFI ranged from 0.002 to −.011. Rutkowski and Svetina (2014) proposed cutoff values for ΔRMSEA and ΔCFI of 0.03 and 0.02, respectively, to achieve metric invariance in large numbers of groups. They also advised against using SRMR in isolation (Ding et al., 2022). The values we obtained are below the cut-off scores suggested by Rutkowski and Svetina (2014). Therefore, we conclude that both models showed measurement invariance among the five English-speaking countries. This means the three-factor structure consisting of three subscales (interest, confidence, and value recognition) was the same among countries under the same condition (language). However, the CFI values of Model A for the MGCFA ranged from 0.876 to 0.889, and the CFI value of the metric invariance of Model C was 0.907. Therefore, from these CFI values, it can be inferred that the metric invariance of Model C based on the EFA results was fitted better than all models of Model A. Thus, it was confirmed that Model C has better measurement invariance than Model A.

Discussion

This study investigated the validity and measurement invariance of the SAM scale from TIMSS 2019 for five English-speaking countries. TIMSS 2019 reported that the validity of the SAM scale for all data was achieved using PCA. However, the measuring tool in international comparative research, such as TIMSS or PISA, has translation effects. This phenomenon threatens the validity of comparative studies between countries. Many researchers have investigated the translation effects caused by different languages using differential item functioning (DIF), MGCFA, and factor analyses (Asil & Gelbal, 2012; El Masri & Andrich, 2020; Gökçe et al., 2021; Oliveri & Ercikan, 2011). Asil and Gelbal (2012) proclaimed that increasing the number of DIF items was caused by linguistic and cultural differences. Gökçe et al. (2021) suggested that research results using TIMSS data should reflect culture, language, curriculum, or other differences. Therefore, this study selected five English-speaking countries as samples to prevent or control translation effects and investigated the validity of the SAM scale.

TIMSS is the widely used data for conducting international comparative research. Checking the validity of measuring tools considering these data characteristics would be a more valuable way to improve the usefulness and accuracy of the scale, rather than discussing data from a few countries of interest or simply combing all the data without considering the diversity among countries. This study investigated the validity of five English-speaking countries from this point of view, providing two meaningful results to be discussed.

Consideration of Two Items for Improving the Accuracy of Measurement

According to the results, two items (19G ("My teacher tells me I am good at mathematics") and 20E ("I would

Table 7. Pattern Matrix of Factor Loadings (Dataset A, PAF).

	Australia			England			Ireland			Singapore			USA		
Item no.	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
16A*	**0.83**	0.08	0.00	**0.82**	0.06	0.00	**0.84**	0.08	−0.02	**0.83**	0.05	0.00	**0.79**	0.10	0.03
16B	**0.47**	0.16	0.13	**0.44**	0.22	0.09	**0.51**	0.14	0.12	**0.46**	0.31	0.05	**0.42**	0.25	0.02
16C	**0.55**	−0.01	0.01	**0.50**	0.01	0.05	**0.55**	0.05	0.01	**0.52**	0.15	0.01	**0.43**	0.13	−0.05
16D*	**0.74**	−0.12	0.11	**0.71**	−0.11	0.12	**0.70**	−0.10	0.12	**0.73**	−0.11	0.08	**0.70**	−0.11	0.15
16E*	**0.84**	0.09	0.00	**0.83**	0.10	−0.01	**0.88**	0.06	−0.04	**0.88**	0.06	−0.03	**0.87**	0.07	−0.01
16F*	**0.74**	0.07	0.02	**0.74**	0.06	−0.04	**0.69**	0.05	0.05	**0.76**	−0.01	−0.02	**0.83**	−0.04	−0.02
16G*	**0.75**	0.15	−0.01	**0.79**	0.08	−0.01	**0.69**	0.13	0.03	**0.81**	0.09	−0.02	**0.88**	0.02	−0.02
16H*	**0.87**	−0.04	−0.04	**0.86**	−0.05	−0.05	**0.84**	−0.04	−0.05	**0.85**	−0.08	−0.04	**0.87**	−0.04	−0.01
16I*	**0.76**	0.16	−0.03	**0.74**	0.15	−0.03	**0.79**	0.15	−0.09	**0.76**	0.21	−0.04	**0.78**	0.16	−0.03
19A*	0.20	**0.59**	0.07	0.28	**0.50**	0.07	0.18	**0.59**	0.07	0.36	**0.51**	0.05	0.27	**0.50**	0.14
19B	−0.12	**0.77**	0.02	−0.09	**0.70**	0.03	−0.11	**0.80**	0.02	−0.02	**0.73**	0.01	−0.09	**0.78**	0.04
19C	0.05	**0.71**	0.01	−0.01	**0.60**	0.09	0.09	**0.63**	−0.03	0.11	**0.78**	0.03	0.11	**0.62**	0.00
19D*	0.20	**0.62**	0.04	0.24	**0.52**	0.06	0.18	**0.58**	0.09	0.39	**0.41**	0.06	0.30	**0.49**	0.12
19E	−0.05	**0.54**	−0.02	0.00	**0.48**	−0.05	−0.04	**0.56**	−0.05	−0.04	**0.60**	−0.02	−0.04	**0.60**	−0.05
19F*[b]	0.21	**0.61**	0.05	0.24	**0.55**	0.03	0.22	**0.55**	0.06	0.42	0.39	0.06	0.31	**0.46**	0.12
19G*[a,b]	0.27	0.26	0.08	0.32	0.12	0.06	0.29	0.19	0.13	0.41	0.21	0.05	0.32	0.19	0.15
19H	0.02	**0.79**	−0.01	−0.01	**0.79**	−0.01	0.02	**0.77**	0.00	0.03	**0.81**	0.01	0.02	**0.79**	0.00
19I	0.11	**0.68**	−0.03	0.13	**0.64**	−0.05	0.07	**0.68**	0.04	0.04	**0.73**	0.02	0.07	**0.68**	−0.03
20A*	0.20	−0.08	**0.61**	0.29	−0.12	**0.51**	0.24	−0.10	**0.58**	0.39	−0.10	**0.42**	0.26	−0.10	**0.57**
20B*	0.20	−0.10	**0.58**	0.21	−0.09	**0.57**	0.21	−0.07	**0.54**	0.29	−0.14	**0.43**	0.20	−0.08	**0.55**
20C*	−0.06	0.06	**0.76**	−0.09	0.01	**0.74**	−0.10	0.02	**0.76**	−0.04	0.09	**0.68**	−0.10	0.03	**0.77**
20D*	−0.06	0.02	**0.78**	−0.03	0.02	**0.76**	−0.08	0.04	**0.78**	−0.01	0.03	**0.73**	−0.02	0.00	**0.75**
20E*[a,b]	0.28	0.14	**0.41**	0.30	0.15	0.39	0.33	0.13	0.36	**0.47**	0.13	0.28	0.38	0.05	0.37
20F*	0.03	−0.01	**0.78**	0.17	−0.05	**0.69**	0.11	−0.03	**0.69**	0.14	−0.04	**0.70**	0.10	−0.05	**0.75**
20G*	−0.08	0.00	**0.83**	0.00	−0.02	**0.80**	−0.08	0.04	**0.77**	−0.05	0.01	**0.80**	−0.08	0.03	**0.83**
20H*	−0.08	0.05	**0.59**	−0.12	0.08	**0.63**	−0.05	0.03	**0.53**	−0.12	0.00	**0.63**	−0.10	0.05	**0.64**
20I*	0.06	0.00	**0.74**	−0.05	0.06	**0.75**	0.04	0.00	**0.70**	0.01	0.04	**0.70**	−0.01	0.03	**0.73**

Note. F1 = interest; F2 = confidence; F3 = value recognition; * = reversely coded items; bold = factor loadings > .4; [a] = factor loadings < 0.4 in more than four countries; [b] = wrongly assigned item.

Table 8. Descriptive Statistics of the SAM Scale (Dataset B).

Item no.	Item statement	Australia (*N* = 4,094)		England (*N* = 1,394)		Ireland (*N* = 1,861)		Singapore (*N* = 2,375)		USA (*N* = 3,574)	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Interest											
16A*	I enjoy learning mathematics	2.79	0.98	2.77	0.93	2.77	0.99	3.08	0.90	2.83	1.00
16B	I wish I did not have to study mathematics	2.72	1.13	2.75	1.13	2.61	1.17	2.74	1.06	2.58	1.15
16C	Mathematics is boring	2.48	1.22	2.43	1.30	2.38	1.19	2.73	1.01	2.50	1.29
16D*	I learn many interesting things in mathematics	2.81	0.92	2.71	0.93	2.70	0.97	3.03	0.84	2.86	0.96
16E*	I like mathematics	2.70	1.01	2.67	1.00	2.70	1.05	2.94	0.96	2.73	1.06
16F*	I like any schoolwork that involves numbers	2.33	0.93	2.28	0.94	2.36	1.00	2.52	0.90	2.35	0.98
16G*	I like to solve mathematics problems	2.55	1.01	2.48	1.00	2.48	1.06	2.76	0.98	2.58	1.05
16H*	I look forward to mathematics class	2.34	0.97	2.30	0.97	2.25	0.99	2.62	0.94	2.47	1.05
16I*	Mathematics is one of my favorite subjects	2.31	1.14	2.14	1.08	2.23	1.15	2.68	1.12	2.49	1.18
Confidence											
19A*	I usually do well in mathematics	3.02	0.88	3.02	0.82	3.00	0.87	2.70	1.00	3.18	0.89
19B	Mathematics is more difficult for me than for many of my classmates	2.77	1.07	2.74	1.15	2.75	1.06	2.60	0.97	2.77	1.09
19C	Mathematics is not one of my strengths	2.60	1.24	2.53	1.33	2.53	1.25	2.45	1.12	2.67	1.34
19D*	I learn things quickly in mathematics	2.73	0.93	2.74	0.88	2.70	0.96	2.72	0.92	2.83	0.96
19E	Mathematics makes me nervous	2.83	1.04	3.02	1.16	2.89	1.11	2.36	1.00	2.80	1.09
19F*	I am good at working out difficult mathematics problems	2.56	0.94	2.64	0.91	2.49	0.95	2.45	0.93	2.65	0.99
19G*	My teacher tells me I am good at mathematics	2.52	0.97	2.57	0.99	2.62	0.99	2.33	0.93	2.68	1.03
19H	Mathematics is harder for me than any other subject	2.82	1.09	2.86	1.20	2.84	1.14	2.72	1.09	2.79	1.18
19I	Mathematics makes me confused	2.53	1.03	2.50	1.03	2.42	1.05	2.41	1.02	2.53	1.10
Value Recognition											
20A*	I think learning mathematics will help me in my daily life	3.24	0.85	3.14	0.87	3.01	0.98	3.15	0.83	3.14	0.91
20B*	I need mathematics to learn other school subjects	3.08	0.85	3.14	0.83	2.98	0.90	3.00	0.82	3.01	0.90
20C*	I need to do well in mathematics to get into the <university> of my choice	3.22	0.90	3.47	0.75	3.35	0.85	3.30	0.78	3.46	0.78
20D*	I need to do well in mathematics to get the job I want	3.15	0.92	3.23	0.90	3.14	0.96	3.20	0.83	3.20	0.91
20E*	I would like a job that involves using mathematics	2.43	1.04	2.31	1.04	2.30	1.05	2.48	0.97	2.41	1.08
20F*	It is important to learn about mathematics to get ahead in the world	3.19	0.85	3.13	0.86	3.02	0.94	3.21	0.77	3.19	0.87
20G*	Learning mathematics will give me more job opportunities when I am an adult	3.45	0.74	3.46	0.71	3.40	0.78	3.39	0.70	3.41	0.78
20H*	My parents think that it is important that I do well in mathematics	3.57	0.67	3.58	0.66	3.61	0.66	3.53	0.65	3.57	0.71
20I*	It is important to do well in mathematics	3.50	0.70	3.57	0.64	3.49	0.74	3.56	0.64	3.56	0.69

Note. *= reversely coded items.

Table 9. CFA Model Comparisons by the Model Fit Indices (Dataset B).

	χ^2					χ^2/df					RMSEA					SRMR					CFI				
Model	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Model A	7,660.179	2,991.964	3,594.105	5,346.854	7,030.453	23.863	9.321	11.197	16.657	21.902	0.068	0.070	0.068	0.071	0.068	0.066	0.073	0.066	0.080	0.069	0.903	0.877	0.901	0.890	0.899
Model B	23,305.080	7,191.366	9,803.296	13,013.634	20,584.585	71.929	22.196	30.257	40.166	63.533	0.119	0.111	0.115	0.114	0.117	0.116	0.117	0.113	0.115	0.122	0.696	0.685	0.713	0.716	0.698
Model C	5,840.119	2,366.465	2,808.915	4,021.872	5,511.687	21.471	8.700	10.327	14.786	20.264	**0.063**	**0.066**	**0.064**	**0.066**	**0.064**	**0.051**	**0.058**	**0.050**	**0.059**	**0.057**	**0.922**	**0.898**	**0.919**	**0.912**	**0.917**
Model D				6,792.223					21.160					0.081					0.097					0.856	

Note. Country: 1 = Australia, 2 = England, 3 = Ireland, 4 = Singapore, 5 = USA. Model A = 3-factor CFA based on the TIMSS 2019 Technical Report, Model B = 1-factor CFA often used in research, Model C = the 3-factor CFA based on EFA results, with two deleted items (19G, 20E), Model D (applied only to Singapore) = a three-factor CFA based on EFA results, which moved three items (19F, 19G, 20E) from each original scale to interest.

Table 10. Model Fit Indices of MGCFA for the Measurement Invariance of the SAM scale (Dataset B).

Model	Level invariance	χ^2 *(df)*	$\Delta\chi^2$ Δ*(df)*	*p*	RMSEA	SRMR	CFI	AIC	BIC	ΔRMSEA	ΔSRMR	ΔCFI
Model A	Configural	26,633.73(1,605)			0.077	0.068	0.889	787,426.46	790,574.51			
	Metric	27,250.36(1,701)	572.71(96)	<.001	0.075	0.071	0.887	787,851.09	790,279.59	0.002	−0.003	0.002
	Scalar	29,834.76(1,797)	2512.51(96)	<.001	0.077	0.073	0.876	790,243.49	791,952.44	−0.002	−0.002	0.011
Model C	Configural	20,556.83(1,360)			0.072	0.053	0.910*	732,003.97	734,932.44			
	Metric	21,150.77(1,448)	547.14(88)	<.001	0.071	0.057	0.907*	732,421.92	734,689.60	0.001	−0.004	0.003
	Scalar	23,580.78(1,536)	2,364.64(88)	<.001	0.073	0.059	0.896	734,675.92	736,282.83	−0.002	−0.002	0.011

Note. Configural: the same items belong to the same construct/metric: equal factor loadings/scalar: equal factor loading and equal intercepts.
*= CFI>.9.

like a job that involves using mathematics")) out of 27 were not appropriate for the SAM scale when using EFA methods controlling for translation effects. The communalities and the loading values of item 19G in the factor TIMSS intended were below 0.40 among all countries. Looking at the meaning of item 19G, the statement, "My teacher tells me I am good at mathematics," refers to a teacher's opinion rather than the student's thoughts. Reynolds et al. (2022) also discussed that the measurement invariance was unsatisfactory with item 19G. Furthermore, the loading values of item 20E were below 0.40 in the factor intended by TIMSS in four countries, and the loading value of this item was above 0.40 in interest factor only in Singapore. Looking at the meaning of item 20E, the statement, "I would like a job that involves using mathematics," is closer to interest than value recognition, and there may be different interpretations due to the cultural differences between Singapore and the four Western countries based on the results of Singapore.

Moreover, it was confirmed that the model based on EFA results with the two items deleted fitted better than the others with CFA results. Also, Model C based on EFA results fitted better than Model A suggested by TIMSS, when using the MGCFA method for measurement invariance testing. Thus, we concluded that it is better to delete or revise the two items (19G and 20E) to improve the measurement accuracy.

The Different Results for the Three Items Only in Singapore

The other significant result is that three items (19F ("I am good at working out difficult mathematics problems"), 19G ("My teacher tells me I am good at mathematics"), and 20E ("I would like a job that involves using mathematics")) were assigned incorrectly to another factor, and the incorrectly assigned loading values of these three items (19F, 19G, and 20E) were above 0.40 in interest factor only in Singapore, even though it was linguistically the same condition as the other four countries. This is presumed to result from cultural differences for the same reason as mentioned above. Although Singapore is a multilingual country, all Singaporean students completed the questionnaire in English. In contrast, 2% of Irish students used questionnaires written in a second language (See Appendix 3). Thus, we could not conclude that the characteristics were due to differences between multilingual and English-speaking countries, even if the ratio was small. The meaning of the items could be interpreted differently depending on country-specific cultural factors rather than on reading ability, even if the items were written in the same language. Directly comparing the results from different countries poses limitations, as pointed out by Schmitt and Allik (2005) and Lindwall et al. (2012). Therefore, future research should investigate results that reflect differences according to cultural background.

Conclusion

Many researchers have used non-cognitive scales in TIMSS to study students' learning attitudes and the relationship between achievement and attitudes in mathematics and science. However, few studies have been conducted to confirm validity measures, such as the items themselves, factor structure, and measurement invariance. Measurement invariance is especially important for analyzing cohort trends in longitudinal and large-scale educational studies (Wurster, 2022). Therefore, it is meaningful to investigate the validity and factor structure of the SAM scales because it has been modified, deleted, and added to over the decades.

This study is significant for testing the validity of survey tools under the same conditions, such as language or culture, before comparing data among various countries in international comparative research. The SAM scale is widely used by many researchers and countries globally; thus, it is meaningful because it provides researchers with a perspective on how to use and deal with the scale before using the data to conduct cross-country comparative studies.

However, this research is limited because it was conducted only in English-speaking countries. It would be good to conduct a validity study more deeply by finding and reflecting certain similar conditions, such as other languages or cultures among the researched countries. In addition, further research using a dataset from Singapore is needed to examine why different results were produced despite the same language conditions to determine whether the results were due to a different culture. Finally, as the result of the statistical analysis cannot be said to reflect all the contexts of the survey, a follow-up study is needed to further secure the validity of the scales by reflecting the analysis of content experts.

Author Note

This work is developed from the paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, United States.

ORCID iDs

Hun Won Choi https://orcid.org/0009-0004-0977-0164
Youn-Jeng Choi https://orcid.org/0000-0001-9803-2681

Author Contributions

Writing—original draft—Hun Won Choi.
Writing—review & editing—Youn-Jeng Choi.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The TIMSS data can be downloaded from http://timssandpirls.bc.edu/ (TIMSS) for free.

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

Aditomo, A., & Klieme, E. (2020). Forms of inquiry-based science instruction and their relations with learning outcomes: Evidence from high and low-performing education systems. *International Journal of Science Education*, *42*(4), 504–525. https://doi.org/10.1080/09500693.2020.1716093

Ardic, E. O., & Gelbal, S. (2017). Cross-group equivalence of interest and motivation items in PISA 2012 Turkey sample. *Eurasian Journal of Educational Research*, *17*(68), 223–238. https://doi.org/10.14689/ejer.2017.68.12

Asil, M., & Gelbal, S. (2012). Cross-cultural equivalence of the PISA student questionnaire. *Egitim Ve Bilim-Education and Science*, *37*(166), 236–249.

Ayob, A., & Yassin, R. M. (2017). A confirmatory factor analysis of the attitude towards mathematics scale using multiply imputed datasets. *International Journal of Advanced and Applied Sciences*, *4*(3), 7–12. https://doi.org/10.6007/IJARBSS/v7-i4/2969

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74–94. https://doi.org/10.1177/009207038801600107

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*(3), 303–316. https://doi.org/10.1177/0049124189017003004

Bolt, D., Wang, Y. C., Meyer, R. H., & Pier, L. (2020). An IRT mixture model for rating scale confusion associated with negatively worded items in measures of social-emotional learning. *Applied Measurement in Education*, *33*(4), 331–348. https://doi.org/10.1080/08957347.2020.1789140

Bulut, H. C. (2021). Item wording effects in psychological measures: Do early literacy skills matter? *Journal of Measurement and Evaluation in Education and Psychology*, *12*(3), 239–253. https://doi.org/10.21031/epod.944067

Bulut, H. C., & Bulut, O. (2022). Item wording effects in self-report measures and reading achievement: Does removing careless respondents help? *Studies In Educational Evaluation*, *72*, 101126. https://doi.org/10.1016/j.stueduc.2022.101126

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Chen, M., & Hastedt, D. (2022). The paradoxical relationship between students' non-cognitive factors and mathematics & science achievement using TIMSS 2015 dataset. *Studies In Educational Evaluation*, *73*, 101–145. https://doi.org/10.1016/j.stueduc.2022.101145

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

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*(1), 55–75. https://doi.org/10.1146/annurev-soc-071913-043137

Ding, Y., Yang Hansen, K., & Klapp, A. (2022). Testing measurement invariance of mathematics self-concept and self-efficacy in PISA using MGCFA and the alignment method. *European Journal of Psychology of Education*, *38*, 709–732. https://doi.org/10.1007/s10212-022-00623-y

El Masri, Y. H., & Andrich, D. (2020). The trade-off between model fit, invariance, and validity: The case of PISA science assessments. *Applied Measurement in Education*, *33*(2), 174–188. https://doi.org/10.1080/08957347.2020.1732384

Ertürk, Z., & Oyar, E. (2021). Examining the measurement invariance of TIMSS 2015 mathematics liking scale through different methods. *International Journal of Assessment Tools in Education*, *8*(1), 67–89. https://doi.org/10.21449/ijate.705426

Eser, D. C. (2021). Investigation of measurement invariance according to home resources: TIMSS 2015 mathematical affective characteristics Questionnaire. *International Journal of Assessment Tools in Education*, *8*(3), 633–648. https://doi.org/10.21449/ijate.817168

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.2307/3151312

Glassow, L. N., Rolfe, V., & Hansen, K. Y. (2021). Assessing the comparability of teacher-related constructs in TIMSS 2015 across 46 education systems: An alignment optimization approach. *Educational Assessment Evaluation and Accountability*, *33*, 105–137. https://doi.org/10.1007/s11092-020-09348-2

Gökçe, S., Berberoğlu, G., Wells, C. S., & Sireci, S. G. (2021). Linguistic distance and translation differential item functioning on trends in International Mathematics and Science

Study Mathematics Assessment Items. *Journal of Psychoeducational Assessment*, *39*(6), 728–745. https://doi.org/10.1177/07342829211010537

Gorsuch, R. (Ed.). (1983). *Factor analysis* (2nd ed.). Lawrence Erlbaum Associates.

Greensfeld, H., & Deutsch, Z. (2022). Mathematical challenges and the positive emotions they engender. *Mathematics Education Research Journal*, *34*, 15–36. https://doi.org/10.1007/s13394-020-00330-1

Guo, X., Hao, X., Deng, W., Ji, X., Xiang, S., & Hu, W. (2022). The relationship between epistemological beliefs, reflective thinking, and science identity: A structural equation modeling analysis. *International Journal of Stem Education*, *9*, 40. https://doi.org/10.1186/s40594-022-00355-x

Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (Eds.). (2006). *Multivariate data analysis* (6th ed.). Pearson Prentice Hall.

Hirsch, O., Chavanon, M., Riechmann, E., & Christiansen, H. (2018). Emotional dysregulation is a primary symptom in adult attention-deficit/hyperactivity disorder (ADHD). *Journal of Affective Disorders*, *232*, 41–47. https://doi.org/10.1016/j.jad.2018.02.007

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., & 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

Kam, C. C. S., & Meyer, J. P. (2023). Testing the nonlinearity assumption underlying the use of reverse-keyed items: A logical response perspective. *Assessment*, *30*(5), 1569–1589. https://doi.org/10.1177/10731911221106775

Lei, P. W., & Wu, Q. (2015). Estimation in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 164–180). Guilford Press.

Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment*, *94*(2), 196–204. https://doi.org/10.1080/00223891.2011.645936

Liou, P. Y., & Lin, J. J. H. (2021). Comparisons of science motivational beliefs of adolescents in Taiwan, Australia, and the United States: Assessing the measurement invariance across countries and genders. *Frontiers in Psychology*, *12*, 674–902. https://doi.org/10.3389/fpsyg.2021.674902

Liu, S., & Meng, L. (2010). Re-examining factor structure of the attitudinal items from TIMSS 2003 in cross-cultural study of mathematics self-concept. *Educational Psychologist*, *30*(6), 699–712. https://doi.org/10.1080/01443410.2010.501102

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., Xu, M. K., Nagengast, B., & Parker, P. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Education & Psychology*, *105*(1), 108–128. https://doi.org/10.1037/a0029907

Martin, M. O., von Davier, M., & Mullis, I. V. S. (Eds.). (2020). *Methods and procedures: TIMSS 2019 technical report*. Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/methods/

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *E-Journal of Applied Psychology*, *93*(3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Michaelides, M. P. (2019). Negative keying effects in the factor structure of TIMSS 2011 motivation scales and associations with reading achievement. *Applied Measurement in Education*, *32*(4), 365–378. https://doi.org/10.1080/08957347.2019.1660349

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge/Taylor & Francis Group.

Mullis, I. V. S., & Martin, M. O. (Eds.) (2017). *TIMSS 2019 assessment frameworks*. Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/frameworks/

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/international-results/

Nunnally, J. C., & Bernstein, I. H. (Eds.). (1994). *Psychometric theory* (3rd ed.). McGraw-Hill, Inc.

Oliveri, M. E., & Ercikan, K. (2011). Do different approaches to examining construct comparability in multilanguage assessments lead to similar conclusions? *Applied Measurement in Education*, *24*(4), 349–366. https://doi.org/10.1080/08957347.2011.607063

Oon, P. T., & Fan, X. (2017). Rasch analysis for psychometric improvement of science attitude rating scales. *International Journal of Science Education*, *39*(6), 683–700. https://doi.org/10.1080/09500693.2017.1299951

Pedhazur, E. J., & Schmelkin, L. P. (Eds.). (1991). *Measurement, design, and analysis: An integrated approach* (student ed.). Lawrence Erlbaum Associates, Inc.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Positive emotions in education. In E. Frydenberg (Ed.), *Beyond coping: Meeting goals, visions, and challenges* (pp. 149–173). Oxford University Press.

Reynolds, K., Khorramdel, L., & von Davier, M. (2022). Can students' attitudes towards mathematics and science be compared across countries? Evidence from measurement invariance modeling in TIMSS 2019. *Studies In Educational Evaluation*, *74*, 101–169. https://doi.org/10.1016/j.stueduc.2022.101169

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale

international surveys. *Educational and Psychological Measurement*, *74*(1), 31–57. https://doi.org/10.1177/0013164413498257

Sabah, S., Hammouri, H., & Akour, M. (2013). Validation of a scale of attitudes toward science across countries using Rasch Model: Findings from TIMSS. *Journal of Baltic Science Education*, *12*(5), 692–702. https://doi.org/10.33225/jbse/13.12.692

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage Publications, Inc.

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg self-esteem scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology*, *89*(4), 623–642. https://doi.org/10.1037/0022-3514.89.4.623

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using M plus and the lavaan/semtools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776

Tabachnick, B. G., & Fidell, L. S. (Eds.). (2001). *Using multivariate statistics* (5th ed.). Needham Heights.

Uyar. (2021). Factor structure and measurement invariance of the TIMSS 2015 mathematics attitude questionnaire: Exploratory structural equation modelling approach. *International Journal of Assessment Tools in Education*, *8*(4), 855–871. https://doi.org/10.21449/ijate.796862

Vaculíková, J., Kočvarová, I., Kalenda, J., Neupauer, Z., Vukčević, M. C., & Włoch, A. (2022). Factor structure of the self-regulation questionnaire among adult learners from Poland, Serbia, Slovakia, and the Czech Republic. *Psicologia*, *35*, 40. https://doi.org/10.1186/s41155-022-00241-z

West, S. G., Taylor, A. B., & Wu, W. (2015). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). Guilford Press.

Willmer, M., Jacobson, J. W., & Lindberg, M. (2019). Exploratory and confirmatory factor analysis of the 9-item Utrecht work engagement scale in a multi-occupational female sample: A cross-sectional study. *Frontiers in Psychology*, *10*, 2771. https://doi.org/10.3389/fpsyg.2019.02771

Wurster, S. (2022). Measurement invariance of non-cognitive measures in TIMSS across countries and across time. An application and comparison of multigroup confirmatory factor analysis, Bayesian approximate measurement invariance and alignment optimization approach. *Studies In Educational Evaluation*, *73*, 101–143. https://doi.org/10.1016/j.stueduc.2022.101143

Appendix 1. Trend of Changes in Items About the SAM Scale in TIMSS.

First item statement	1995		1999		2003		2007		2011		2015		2019		Final No.	Final item statement
Interest																*Interest*
How much do you like mathematics? (different likert scales)	21	o	21	o		x										
I enjoy learning mathematics	23	o	24	o	8	o	8	o	14	o	17	o	16	o	16a	I enjoy learning mathematics
Mathematics is boring	23	o	24	o		x	8	o	14	o	17	o	16	o	16c	Mathematics is boring
Mathematics is easy subject	23	o	24	o		x										
Mathematics is important to everyone's life	23	o	24	o		x										
I would like a job that involved using mathematics	23	o	24	o		x										
I would like to take more mathematics in school					8	o	8	o		x						
Sometimes, when I do not initially understand a new topic in mathematics, I know that I will never really understand it					8	o		x								
I like mathematics							8	o	14	o	17	o	16	o	16e	I like mathematics
I wish I did not have to study mathematics									14	o	17	o	16	o	16b	I wish I did not have to study mathematics
I learn many interesting things in mathematics									14	o	17	o	16	o	16d	I learn many interesting things in mathematics
I like any schoolwork that involves numbers											17	o	16	o	16f	I like any schoolwork that involves numbers
I like to solve mathematics problems											17	o	16	o	16g	I like to solve mathematics problems
I look forward to mathematics class											17	o	16	o	16h	I look forward to mathematics class
Mathematics is one of my favorite subjects											17	o	16	o	16i	Mathematics is one of my favorite subjects
Confidence																*Confidence*
I usually do well in mathematics.	17	o	16	o	8	o	8	o	16	o	19	o	19	o	19a	I usually do well in mathematics.
I would like mathematics much more if it were not so difficult.			17	o		x										
Although I do my best, mathematics is more difficult for me than for many of my classmates			17	o	8	o(R)	8	o	16	o	19	o	19	o	19b	Mathematics is more difficult for me than for many of my classmates
Nobody can be good in every subject, and I am just not talented in mathematics.			17	o		x										
Sometimes, when I do not understand a new topic in mathematics initially, I know that I will never really understand it			17	o		x										
Mathematics is not one of my strengths			17	o	8	o	8	o	16	o	19	o	19	o	19c	Mathematics is not one of my strengths
I learn things quickly in mathematics					8	o	8	o	16	o	19	o	19	o	19d	I learn things quickly in mathematics
Mathematics makes me confused and nervous									16	o(M)	19	o	19	o	19e	Mathematics makes me nervous

(continued)

Appendix 1. (continued)

First item statement	1995		1999		2003		2007		2011		2015		2019		Final No.	Final item statement
I am good at working out difficult mathematics problems									16	o	19	o	19	o	19f	I am good at working out difficult mathematics problems
My teacher thinks I can do well in mathematics <programs/classes/ lessons> with difficult materials									16	o		x		x		
My teacher tells me I am good at mathematics									16	o	19	o	19	o	19g	My teacher tells me I am good at mathematics
Mathematics is harder for me than any other subject									16	o	19	o	19	o	19h	Mathematics is harder for me than any other subject
Mathematics makes me confused									16	o(M)	19	o	19	o	19i	Mathematics makes me confused
Value recognition																*Value recognition*
I need to do well in mathematics to get the job I want	24	o	25	o	9	o	9	o	16	o	20	o	20	o	20d	I need to do well in mathematics to get the job I want
I need to do well in mathematics to please my parent(s)	24	o	25	o		x										
I need to do well in mathematics to get into the <secondary school> or university I prefer	24	o	25	o	9	o(R)	9	o	16	o	20	o	20	o	20c	I need to do well in mathematics to get into the university of my choice
I need to do well in mathematics to please myself	24	o	25	o		x										
I think learning mathematics will help me in my daily life					9	o	9	o	16	o	20	o	20	o	20a	I think learning mathematics will help me in my daily life
I need mathematics to learn other school subjects					9	o	9	o	16	o	20	o	20	o	20b	I need mathematics to learn other school subjects
I would like a job that involved using mathematics					9	o		x	16	o(R)	20	o	20	o	20e	I would like a job that involves using mathematics
It is important to learn about mathematics to get ahead in the world											20	o	20	o	20f	It is important to learn about mathematics to get ahead in the world
Learning mathematics will give me more job opportunities when I am an adult											20	o	20	o	20g	Learning mathematics will give me more job opportunities when I am an adult
My mother thinks it is important for me to do well in mathematics at school.	13	o	12	o		x					20	o(R)	20	o	20h	My parents think that it is important that I do well in mathematics
I think it is important to do well in mathematics at school.	16	o	15	o		x			14	o(R)	20	o	20	o	20i	It is important to do well in mathematics
Most of my friends think it is important to do well in mathematics at school.	15	o	14	o		x										

Note. R = revised item; M = Items were merged in 2011 and they were separated in 2015.

Appendix 2. Average Mathematics Achievement of Eighth-Grade Students.

Country	*M* (SE)
Singapore	**616 (4.0)**
Chinese Taipei	612 (2.7)
Rep. of Korea	607 (2.8)
Japan	594 (2.7)
Hong Kong SAR	578 (4.1)
Russian Federation	543 (4.5)
Ireland	**524 (2.6)**
Lithuania	520 (2.9)
Israel	519 (4.3)
Australia	**517 (3.8)**
Hungary	517 (2.9)
The United States	**515 (4.8)**
England	**515 (5.3)**
Finland	509 (2.6)
Norway	503 (2.4)
Sweden	503 (2.5)
Cyprus	501 (1.6)
Portugal	500 (3.2)
TIMSS Scale center point	500
Italia	497 (2.7)
Turkiye	496 (4.3)
Kazakhstan	488 (3.3)
France	483 (2.5)
New Zealand*	482 (3.3)
Bahrain	481 (1.7)
Romania	479 (4.3)
United Arab Emirates	473 (1.9)
Georgia	461 (4.3)
Malaysia	461 (3.2)
Iran	446 (3.7)
Qatar	443 (4.0)
Chile	441 (2.8)
Lebanon	429 (2.9)
Jordan	420 (4.3)
Egypt	413 (5.2)
Oman	411 (2.8)
Kuwait	403 (5.0)
Saudi Arabia	394 (2.5)
South Africa	389 (2.3)
Morocco	388 (2.3)

Note. Bold = selected countries.

*Met guidelines for sample participation rates only after replacement schools were included.

Appendix 3. Language Usage for the Context Questionnaire on Mathematics.

	Australia		England		Ireland		Singapore		USA	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%
English (first language)	8,935	99	3,209	95	3,969	96	4,852	100	8,460	97
Second language	–	–	–	–	97	2	–	–	–	–
Missing	125	1	156	5	52	1	1	0	238	3
Total	9,060	100	3,365	100	4,118	100	4,853	100	8,698	100

Appendix 4. Structure Matrix of Factor Loadings (Dataset A, PAF).

	Australia			England			Ireland			Singapore			USA		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
16A*	**0.88**	**0.57**	**0.46**	**0.85**	**0.49**	0.37	**0.87**	**0.61**	**0.41**	**0.86**	**0.54**	0.37	**0.87**	**0.57**	**0.42**
16B	**0.68**	**0.51**	**0.42**	**0.67**	**0.49**	0.36	**0.69**	**0.52**	**0.40**	**0.68**	**0.58**	0.31	**0.61**	**0.51**	0.29
16C	**0.71**	**0.46**	0.36	**0.71**	**0.44**	0.30	**0.73**	**0.50**	0.35	**0.66**	**0.48**	0.29	**0.65**	**0.50**	0.27
16D*	**0.71**	0.35	**0.48**	**0.70**	0.29	0.39	**0.72**	0.39	**0.45**	**0.70**	0.32	**0.40**	**0.71**	0.33	**0.46**
16E*	**0.90**	**0.59**	**0.46**	**0.89**	**0.52**	0.38	**0.90**	**0.62**	**0.42**	**0.89**	**0.57**	0.37	**0.91**	**0.58**	**0.41**
16F*	**0.79**	**0.51**	**0.42**	**0.76**	**0.43**	0.30	**0.75**	**0.52**	**0.40**	**0.74**	**0.44**	0.33	**0.80**	**0.46**	0.37
16G*	**0.84**	**0.58**	**0.44**	**0.83**	**0.48**	0.35	**0.79**	**0.58**	0.39	**0.84**	**0.55**	0.36	**0.89**	**0.54**	**0.41**
16H*	**0.82**	**0.47**	**0.40**	**0.81**	0.39	0.33	**0.80**	**0.49**	0.35	**0.79**	**0.41**	0.34	**0.84**	**0.48**	0.39
16I*	**0.84**	**0.60**	**0.40**	**0.82**	**0.52**	0.33	**0.84**	**0.63**	0.36	**0.87**	**0.65**	0.34	**0.85**	**0.62**	0.36
19A*	**0.61**	**0.71**	0.36	**0.60**	**0.62**	0.31	**0.59**	**0.73**	0.31	**0.66**	**0.73**	0.26	**0.61**	**0.68**	0.34
19B	0.39	**0.76**	0.18	0.36	**0.74**	0.11	**0.41**	**0.77**	0.17	**0.42**	**0.76**	0.11	**0.41**	**0.78**	0.14
19C	**0.57**	**0.84**	0.27	**0.52**	**0.83**	0.22	**0.53**	**0.78**	0.23	**0.58**	**0.86**	0.18	**0.56**	**0.83**	0.20
19D*	**0.61**	**0.73**	0.36	**0.59**	**0.63**	0.27	**0.60**	**0.75**	0.31	**0.64**	**0.64**	0.29	**0.63**	**0.70**	0.33
19E	0.31	**0.58**	0.10	0.26	**0.57**	0.07	0.32	**0.55**	0.07	0.32	**0.60**	0.05	0.32	**0.60**	0.06
19F*	**0.62**	**0.73**	0.36	**0.60**	**0.62**	0.27	**0.60**	**0.72**	0.31	**0.66**	**0.65**	0.27	**0.63**	**0.66**	0.34
19G*	**0.47**	**0.43**	0.31	**0.46**	0.29	0.25	**0.45**	**0.41**	0.28	**0.54**	**0.46**	0.25	**0.50**	**0.42**	0.32
19H	**0.50**	**0.81**	0.22	**0.45**	**0.82**	0.15	**0.52**	**0.81**	0.20	**0.49**	**0.83**	0.14	**0.50**	**0.84**	0.16
19I	**0.52**	**0.76**	0.23	**0.50**	**0.76**	0.16	**0.55**	**0.74**	0.24	**0.49**	**0.76**	0.15	**0.49**	**0.78**	0.14
20A*	**0.47**	0.21	0.70	**0.45**	0.13	**0.62**	**0.49**	0.20	**0.65**	**0.50**	0.17	**0.58**	**0.48**	0.17	**0.67**
20B*	**0.44**	0.18	0.67	**0.42**	0.12	**0.63**	**0.43**	0.18	**0.62**	**0.40**	0.08	**0.56**	**0.42**	0.15	**0.64**
20C*	0.37	0.22	0.75	0.22	0.09	**0.69**	0.29	0.16	**0.71**	0.31	0.13	**0.69**	0.29	0.13	**0.72**
20D*	0.37	0.21	0.76	0.30	0.14	**0.73**	0.33	0.19	**0.74**	0.32	0.11	**0.73**	0.35	0.14	**0.73**
20E*	**0.60**	**0.43**	0.60	**0.56**	0.37	**0.57**	**0.59**	**0.45**	**0.57**	**0.65**	**0.44**	**0.52**	**0.59**	0.34	**0.56**
20F*	**0.44**	0.22	0.79	**0.43**	0.15	**0.75**	**0.45**	0.21	**0.75**	**0.42**	0.12	**0.76**	**0.43**	0.15	**0.80**
20G*	0.36	0.19	0.79	0.32	0.13	**0.78**	0.34	0.19	**0.74**	0.31	0.08	**0.77**	0.33	0.14	**0.80**
20H*	0.26	0.16	0.58	0.19	0.11	**0.56**	0.22	0.13	**0.54**	0.17	0.05	**0.55**	0.22	0.10	**0.59**
20I*	**0.43**	0.22	0.77	0.32	0.15	**0.73**	**0.40**	0.22	**0.73**	0.35	0.14	**0.70**	0.36	0.17	**0.74**

Note. * = reversely coded item, bold = factor loadings > 0.4.