European Journal of Contemporary Education
E-ISSN 2305-6746
2022. 11(2): 504-514
DOI: 10.13187/ejced.2022.2.504
https://ejce.cherkasgu.press

Psychometric Properties of Anxiety Towards Mathematics Scale using Samples from Four Continents

Jacob Owusu Sarfo [a,*], Arturo García-Santillán [b], Edward Wilson Ansah [a], Henry Adusei [c], Josephine Cudjoe Sarfo [d], Rafael Valdece Sousa Bastos [e], Philip Soyiri Donyeh [f], Violetta S. Molchanova [g,h], Marina Drushlyak [i], Olena Semenikhina [i], Somayeh Zand [j], Violeta Enea [k], Sadia Malik [l], Farzana Ashraf [m], Najma Iqbal Malik [l], Hattaphan Wongcharee [n], Felix O. Egara [o], Arun Tipandjan [p], Uzma Azam [q], Mohammed Salah Hassan [r], Mai Helmy [s,t], Zahir Vally [u], Reza Najafi [v]

[a] University of Cape Coast, Cape Coast, Ghana
[b] Universidad Cristóbal Colón, Veracruz, México
[c] University of Education, Winneba, Ghana
[d] University of South Africa, Pretoria, South Africa
[e] São Francisco University, Brazil
[f] University of Ghana, Legon, Ghana
[g] Cherkas Global University, Washington, USA
[h] Volgograd State University, Volgograd, Russian Federation
[i] Makarenko Sumy State Pedagogical University, Sumy, Ukraine
[j] University of Milano-Bicocca, Milan, Italy
[k] Alexandru Ioan Cuza University, Iasi, Romania
[l] University of Sargodha, Sargodha, Pakistan
[m] COMSATS University, Lahore, Pakistan
[n] Pathumwan Institute of Technology, Bangkok, Thailand
[o] University of Nigeria, Nsukka, Nigeria
[p] International Centre for Psychological Counseling and Social Research, Puducherry, India
[q] Aligarh Muslim University, Uttar Pradesh, India
[r] University of Malaya, Kuala Lumpur, Malaysia
[s] Sultan Qaboos University, Muscat, Oman
[t] Menoufia University, Shebin El-Kom, Egypt
[u] United Arab Emirates University, Al Ain, United Arab Emirates

[*] Corresponding author
E-mail addresses: jacob.sarfo@ucc.edu.gh (J.O. Sarfo)

[v] University of Padova, Padova, Italy

Abstract

This study aimed to examine the factor structure and psychometric properties of the Anxiety Towards Mathematics Scale across four continents. We adopted and translated the original Spanish version of the 24-item Anxiety Towards Mathematics Scale (ATMS-24; Muñoz, Mato-Vazquez, 2007) to collect 4,338 responses from Egypt, Ghana, India, Iran, Malaysia, Mexico, Nigeria, Pakistan, Romania, Thailand, Ukraine, and United Arab Emirates. Also, we conducted an Exploratory Factor Analysis (EFA) on the ATMS-24 to examine whether the data fit well across cultures. Furthermore, we modified the full-length ATMS-24 to a short form (11-items: ATMS-11) using the Gradual Response Model (GRM) of Item Response Theory (IRT) and further conducted an analysis of measurement invariance. The EFA conducted indicated that the ATMS-24 fit the data well across cultures. The new ATMS-11 version has adequate configural, metric, and scalar invariance in seven countries and the overall sample. The ATMS-11 offers a valid, reliable, and parsimonious means to assess mathematics Anxiety (MA) among students from varied cultures. The factor structure and psychometric properties of ATMS-11 support its use for MA assessment in both male and female students across locations in Africa, Asia, Europe, and South America.

Keywords: anxiety towards mathematics scale, cultures, mathematics anxiety, psychometric properties.

1. Introduction

Mathematics is a universal language and attracts a growing consideration from educational, clinical, social, and personal perspectives (Silver et al., 2021; Waller, Flood, 2016). Several human activities in all cultures require the use of some level of mathematics to undertake either basic or complex tasks (García-Santillán et al., 2016; García-Santillán et al., 2018; Polly et al., 2018). Notwithstanding the cognitive dimension of learning and using mathematics globally, some people experience fear, apprehension, tension, worry, frustration, and dislike due to the emotional dimension of mathematics (Dowker et al., 2016). This negative emotion associated with the teaching, learning, achievement, and application of mathematics in academic, career, and daily life was first termed "number anxiety" by Dreger and Aiken (1957).

Currently, number anxiety is called mathematics anxiety (MA) in present educational and research circles (Dowker et al., 2016). Generally, MA can be described as a state of apprehension associated with executing a mathematical task, exposure to numbers and mathematical operations, and classroom evaluations (García-Santillán et al., 2018). Several factors often link unpleasant cognitive, affective, and behavioural responses toward mathematics-related activities. These factors include gender, age, performance in mathematics, teaching approach, academic stress, motivation, parental influence, and biological causes (Dowker et al., 2016; Jamieson et al., 2016; Keshavarzi, Ahmadi, 2013; Silver et al., 2021). As MA affects all ages and cultures, it is essential to assess it with brief tools that are also valid and reliable (García-Santillán et al., 2016).

Notwithstanding this global need for testing MA, most existing instruments are lengthy (Suinn, Winston, 2003) with limited cross-cultural sensitivity (García-Santillán et al., 2018; Pajares, Urdan, 1996; Mahmood, Khatoon, 2011; Muñoz, Mato, 2007), and gender sensitivity (Bai et al., 2009). Moreover, some of these existing instruments have narrow age ranges (Beasley et al., 2001; Chiu, Henry, 1990) and are often old (Brush, 1978; Hopko, 2003; Suinn et al., 1988); thus, they lost touch with contemporary. Besides, none of these measures of MA had been used across four continents to provide robust cross-cultural psychometric evidence to the best of our knowledge.

Considering the above weaknesses, we sought to explore the cross-cultural value of the original Spanish version of the 24-item Anxiety Towards Mathematics Scale (ATMS-24; Muñoz, Mato-Vazquez, 2007). As a unidimensional instrument, all items in the ATMS-24 measure the same latent construct. Although the ATMS-24 has a good factor structure and psychometric properties, a shorter version of the scale with cross-cultural psychometric evidence is desirable in both research and practice. Furthermore, the Principal Components Analysis was used to extract factors in the original ATMS-24 (García-Santillán et al., 2018) like other similar instruments (Moreno-García et al., 2018; Widaman, 2007). By definition, most psychological constructs like MA are a reflective measure (i.e. the variance in the items are due to a latent trait), making an

Exploratory Factor Analysis the most desirable method of extraction (Ellwart, Konradt, 2011). Consequently, a shortened form of ATMS-24 as a reflective measurement offers a suitable brief tool for assessing MA across cultures (Ellwart, Konradt, 2011; Tetrick, Buffardi, 2006).

2. Method

2.1. Participants

We used a quantitative, non-experimental cross-sectional survey to select 4,339 participants from 12 countries (Mexico, Ghana, Pakistan, Iran, Ukraine, Thailand, Romania, Nigeria, India, Malaysia, Egypt, and the United Arab Emirates). The overall sample included 59.9 % females and 40.1 % males. See Table 1 for more details about sample sociodemographic description according to country and region/location collected.

Table 1. Sample Sociodemographic Description According to Country and Region Collected

Countries	Region/ City	*N*	Age %					Gender %	
			12-15	16-20	21-23	24-30	>30	Male	Female
Mexico	Veracruz	201	3	44.3	52.7	0	0	65.7	34.3
Ghana	Koforidua	164	2.44	70.12	22.56	3.05	1.83	57.3	42.7
	Cape Coast	230	0.4	61.3	38.3	0	0	40.4	59.6
Pakistan	Sargodha	394	0	69.8	29.7	0.5	0	42.6	57.4
	Faisalabad	156	0	48,7	51,3	0	0	25.6	74.4
	Rawalpindi	204	0	62.7	37.3	0	0	40.7	59.3
	Lahore	331	0	56.5	43.5	0	0	44.7	55.3
Iran	Tehran	151	0	66.7	33.3	0	0	52.3	47.7
	Qom	155	0	71.6	28.4	0	0	48.4	51.6
Ukraine	Sumy	101	2	74.2	18.8	5	0	31.7	68.3
Thailand	Bangkok	155	0	39.4	60.6	0	0	52.9	47.1
Romania	(Online)	194	0	100	0	0	0	42.8	57.2
Nigeria	Enugu State	117	0	99.1	0.9	0	0	45.3	54.7
India	Puducherry	250	0.4	57.2	40.4	2	0	18.8	81.2
	Uttar Pradesh	207	0	65.2	34.8	0	0	36.7	63.3
Malaysia	(Online)	511	100	0	0	0	0	54.8	45.2
Egypt	(Online)	501	0	53.1	35.9	10.2	0.8	16.8	83.2
United Arab Emirates	(Online)	317	0	71	29	0	0	28.1	71.9
Total		4339	12.1	56.2	29.9	1.6	0.2	40.1	59.9

2.2. Instruments

The data collection instrument contained the following measures: Part A (ATMS-24, Muñoz, Mato-Vazquez, 2007) and Part B (Demographic details, e.g. age, gender, location). Each item in ATMS-24 was rated on a five-point scale, from Strongly disagree (1) to Strongly agree (5).

2.3. Data Collection

Data was collected by a collaborative international team of scientists across 18 research sites in 12 countries across four continents (listed in the participants' locations). Following a call for collaborators by the first author on www.researchgate.net (a social networking site for scientists and researchers) in October 2019, several collaborators applied to join the cross-cultural project. This data collection represents Phase 1 of the project, which ended in October 2020. Ethical approval for the project was obtained from the Institutional Review Board of the International Network Center for Applied Research (INCFAR-IRB/009/01-2020). Nonetheless, individual collaborators were allowed to apply for local or institutional ethics approval. Furthermore, all collaborators were permitted to translate the study protocol and instrument (where necessary) into their respective languages.

The data collection was conducted face-to-face and using an online questionnaire disseminated through e-mail and on social media platforms (i.e. LinkedIn, Facebook, Twitter, and

Instagram). Each collaborating site was to collect a minimum of 150 students (relaxed in two cases: Ukraine [Sumy] and Nigeria [Enugu State] from their respective communities (relaxed in four cases: Romania, Malaysia, Egypt, and United Arab Emirates, where data collection was conducted using online platforms). Any person between 12 and 18 years old could respond to the questionnaire following parental and individual consent. Participants stated their agreement through an informed consent form that ensured confidentiality about their identity. Moreover, each participant was informed to answer the questionnaire only once.

2.4. Data Analysis

An Exploratory Factor Analysis was conducted using the polychoric correlation matrix, Factor version 10.9.02 (Ferrando, Lorenzo-Seva, 2017). One response was dropped due to an almost full incomplete response. The sample adequacy was accessed through Kaiser-Meyer-Olkin (KMO) and the Bartlett test of sphericity was calculated with Kaiser and Chi-square with its p-value. Concerning the extraction method, we used the Robust Diagonally Weighted Least Squares (RDWLS). With Robust Promin Rotation, the RDWLS is adequate for ordinal and non-normal data (Lorenzo-Seva, Ferrando, 2019).

Additionally, we used a parallel analysis based on the Minimum Rank Factor Analysis with 500 simulations to retain factors (Timmerman, Lorenzo-Seva, 2011). The number of factors to be extracted was based on three criteria: the scree plot, eigenvalue greater than one, and the parallel analysis. Besides, the unidimensionality was assessed with Closeness to Unidimensionality Assessment (Uni-Co; Ferrando, Lorenzo-Seva, 2018), with values greater than 0.95 suggesting that data can be treated as unidimensional. Also, the model fit indices were assessed using the Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and χ^2 mean with variance adjusted. Besides, Cronbach's Alpha, McDonald's Omega, and Greatest Lower Bound were used to estimate the reliability of the data (Woodhouse, Jackson, 1977).

Specifically, data cleaning reduced the responses from 4,339 to 4,306 cases using listwise deletion. To reduce the number of items and to know the parameters of the items, parameters *a* (discrimination/slope) and parameters *b* (difficulty/threshold) were calculated using the Gradual Response Model (GRM) of the Item Response Theory (Samejima, 1969) with mirt package [version 1.30] in R Software (Chalmers, 2012; R Core Team, 2019). We also estimated the scale invariance between males and females by conducting a confirmatory multigroup analysis using the lavaan package in R software (R Core Team, 2019; Rosseel, 2012) with an algorithm from Svetina et al. (2019). Subsequently, we tested three levels of invariance. These levels were configural invariance (which tests whether the factor structure is the same between groups), metric invariance (which tests whether the factor loadings of the items are the same between groups), and scalar invariance (which tests whether the intercepts are the same between groups).

3. Results

The Exploratory Factor Analysis showed that the ATMS-24 had sufficient data adequacy; KMO = 0.97 (CI Bootstrap 95 % 0.972 – 0.972); Bartlett test of sphericity, χ^2 (276; N = 4,338) = 49,449.9, p < .001. Two factors showed eigenvalues greater than one, explaining 59.1% of the variance in the data. The first, second, and third factors presented eigenvalues of 12.6, 1.61, and 0.97, respectively. However, the parallel analysis suggested extracting a single factor, with the explained variance of the second observed factor of 7.10 % and simulated data of 8.03 %. Details of factor loadings can be seen in Table 2.

In addition, another evidence of unidimensionality was based on the UniCo = 0.98 (CI 95 %, 0.974 – 0.983) which was greater than 0.95. For the goodness of fit, the scale produced the following statistics: χ^2 (252, N = 4,338) = 6,570.1, p < .001; CFI = 0.98 (CI 95 %, 0.979 – 0.984); TLI = 0.98 (CI 95 %, 0.981 – 0.986); RMSEA = 0.076 (CI 95 %, 0.071 – 0.079). Furthermore, the reliability estimates of data from the Cronbach's Alpha, McDonald's Omega, and Greatest Lower Bound to Reliability were 0.96, 0.96, and 0.98 respectively.

After identifying the unidimensional structure, we sought to reduce the scale with Item Response Theory using GRM (Samejima, 1969) for the 24 items of the ATMS-24 scale. Parameters *a* (discrimination/slope) and parameters *b* (difficulty/threshold) of the items are shown in Table 3. We decided to retain items with substantial discrimination that covered a large portion of the construct without harming the theoretical contribution from the literature (i.e. without excluding all items from the factors found in other papers). Therefore, items 1, 3, 4, 8, 9, 10, 12, 14, 15, 17, 20, 21, and 22 were excluded from the original ATMS-24 scale. The remaining 11 items which formed the ATMS-11 were included in the subsequent analysis.

Table 2. Factor Loadings of the Anxiety Towards Mathematics Scale based on an Exploratory Factor Analysis

Item	Factor 1	h²
1. I get nervous when I think about the math test the day before.	.73	.74
2. I feel nervous when I get the math test questions.	.76	.81
3. I get nervous when I open the math book and find a page full of problems.	.73	.66
4. I feel nervous thinking about the math test when there is only one hour left to do it.	.70	.64
5. I feel nervous when I think about the math test that I have to do next week.	.74	.74
6. I get nervous when I know that I will still have math classes in the next course.	.78	.73
7. I feel nervous when I think about the math test that I have next week.	.78	.80
8. I get nervous when someone looks at me while I do math homework.	.67	.68
9. I feel nervous when I check a purchased ticket after paying.	.47	.57
10. I feel nervous when I start studying for a math test.	.69	.72
11. The math exams make me nervous.	.78	.83
12. I feel nervous when the teacher leaves me with math problems to solve at home and I have to deliver them the next day in class.	.75	.68
13. It makes me nervous to do mathematical operations.	.77	.68
14. I feel nervous having to explain a math problem to the teacher.	.69	.82
15. I get nervous when I do the final math exam.	.71	.74
16. I feel nervous when they give me a list of math exercises.	.77	.76
17. I feel nervous when I try to understand another classmate explaining a math problem.	.69	.69
18. I feel nervous when I do a math assessment test.	.75	.65
19. I feel nervous when I see/hear my teacher explaining a math problem.	.74	.76
20. I am nervous to receive the final (exam) math grades.	.67	.72
21. I feel nervous when I want to find out about the change in the store.	.59	.75
22. I feel nervous when we get a problem and a partner finishes it before me.	.57	.51
23. I feel nervous when I have to explain a problem in math class.	.74	.74
24. I feel nervous when I start doing my homework.	.78	.82
Eigenvalue	12.1	
% variance explained	57.3	
Cronbach's Alpha	.96	
McDonald's Omega	.96	
Greatest Lower Bound	.98	

Note. N = 4,338; h² = communalities; Extraction method Robust Diagonally Weighted Least Squares with Robust Promin rotation.

Table 3. Items Discrimination and Difficulty Parameters

	a	b_1	b_2	b_3	b_4
1.	1.84	-1.57	.0.55	-0.06	0.96
2.	2.03	-1.71	-0.59	-0.17	1.34
3.	1.97	-1.34	-0.34	0.17	1.23
4.	1.77	-1.71	-0.73	-0.20	1.04

5.	1.98	-1.21	-0.15	0.38	1.38
6.	2.29	-0.99	-0.12	0.32	1.06
7.	2.33	-1.08	-0.18	0.27	1.40
8.	1.62	-0.99	-0.05	0.55	1.46
9.	0.93	-1.43	0.21	1.74	3.45
10.	1.84	-1.16	0.11	0.67	1.73
11.	2.30	-1.30	-0.46	0.10	1.07
12.	2.13	-1.27	-0.23	0.31	1.16
13.	2.33	-1.23	-0.15	0.44	1.60
14.	1.78	-1.57	-0.50	0.13	1.31
15.	1.81	-1.59	-0.59	-0.02	1.16
16.	2.34	-1.32	-0.26	0.37	1.49
17.	1.77	-1.05	0.07	0.64	1.80
18.	2.22	-1.20	-0.25	0.34	1.50
19.	2.00	-0.92	0.18	0.71	1.69
20.	1.61	-1.73	-0.74	-0.15	1.04
21.	1.26	-1.16	-0.07	0.79	2.24
22.	1.24	-1.69	-0.47	0.29	1.76
23.	2.10	-1.41	-0.44	0.06	1.15
24.	2.29	-0.83	0.06	0.52	1.47

Note. a = discrimination parameter. b = threshold (difficulty) parameter. Parameters of discrimination and difficulty estimated by Graded Response Model (Samejima, 1969). N = 4,306.

The analysis of measurement invariance (MI) was conducted using multigroup confirmatory factor analysis. Table 4 presents the comparison between configural, metric, and scalar models of the short version of the original scale, ATMS-11. Following the recommendations of Chen (2007), invariance is established when the CFI does not decrease by 0.010, and the RMSEA does not increase by 0.015 when the loads and intercepts are fixed as compared with the prior model when they were not fixed. We found configural, metric, and scalar invariance in Mexico, Ghana, Pakistan, Iran, Romania, India, Egypt, and the overall sample.

Table 4. Measurement Invariance (MI) Testing between Male and Female According to Country

		χ^2	df	*p* value	RMSEA	CFI	TLI
Mexico	Configural	229.08	88	0	0.127	0.955	0.944
	Metric	254.80	110	0	0.115	0.954	0.954
	Scalar	238.26	120	0	0.100	0.962	0.966
Ghana	Configural	392.72	88	0	0.13	0.948	0.935
	Metric	431.03	110	0	0.12	0.945	0.945
	Scalar	417.82	120	0	0.11	0.949	0.954
Pakistan	Configural	518.57	88	0	0.095	0.971	0.964
	Metric	570.67	110	0	0.088	0.969	0.969
	Scalar	554.20	120	0	0.082	0.971	0.973
Iran	Configural	395.17	88	0	0.152	0.944	0.929
	Metric	433.14	110	0	0.139	0.941	0.941
	Scalar	413.85	120	0	0.127	0.946	0.951

Romania	Configural	280.55	88	0	0.151	0.910	0.887
	Metric	317.04	110	0	0.140	0.903	0.903
	Scalar	305.24	120	0	0.127	0.913	0.920
India	Configural	448.85	88	0	0.134	0.935	0.918
	Metric	484.79	110	0	0.122	0.932	0.932
	Scalar	483.07	120	0	0.115	0.934	0.940
Egypt	Configural	682.13	88	0	0.164	0.972	0.965
	Metric	674.13	110	0	0.143	0.973	0.973
	Scalar	649.62	120	0	0.133	0.975	0.977
Total	Configural	2,954.85	88	0	0.121	0.959	0.949
	Metric	3,156.62	110	0	0.113	0.955	0.955
	Scalar	2,975.84	120	0	0.105	0.958	0.961

Note. MI was not accessed in Malaysia, Nigeria, and United Arab Emirates because some response categories had missing data; MI in Thailand and Ukraine was not accessed because of the number of persons in each group. We adopted Svetina et al.'s (2019) algorithm.

However, MI was not accessed in Malaysia, Nigeria, and the United Arab Emirates because some response categories were empty in some groups. Also, MI in Thailand and Ukraine was not accessed because of the number of responses in some specific categories. As a pre-requisite for the Wu and Estabrook (2016) measurement invariance test for ordered categorical outcomes, we needed to have responses in each category of the item response scale (e.g., 1 – Strongly disagree; 5 – Strongly agree). Furthermore, ATMS-11 had reliability estimates of 0.93, 0.95, and 0.95 for Cronbach's Alpha, McDonald's Omega, and Greatest Lower Bound to Reliability. See Figure 1 for the test information curve.

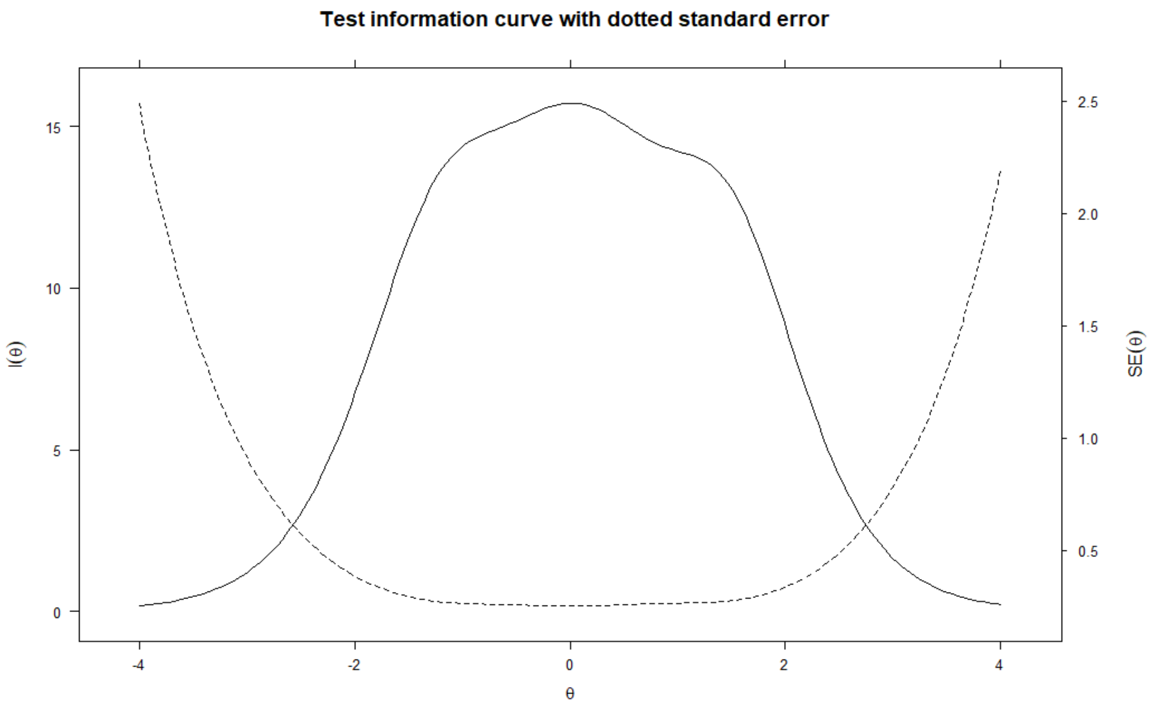

Fig. 1. Test Information Curve
Note. The continuous line represents the test information curve. The dotted line represents the standard error of the measurement.

4. Discussion

Our study explored the psychometric properties of the ATMS-24 (Muñoz, Mato-Vazquez, 2007) and attempted to evaluate the level of MA experience of students across cultures. We further examined the factor structure and psychometric properties of the short form of the original scale, ATMS-11. The present study showed a stable unidimensional structure of the instrument across samples from the four continents. With adequate KMO and Bartlett test of Sphericity, the Exploratory Factor Analysis confirmed the results obtained by García-Santillán et al. (2018) and Muñoz and Mato-Vazquez (2007) among Mexican samples. Significantly high reliability estimates of above 0.90 for the scale's Cronbach's alpha, McDonald's Omega, and Greatest Lower Bound confirmed the scale's reliability (Field, 2009; Hair et al., 1998).

Per previous literature (Samejima, 1969), we were able to reduce the original ATMS from 24 to 11 items. A shorter scale allows future researchers and practitioners to apply the instrument without harming the participants' responses (e.g., acquiescent responses, fatigue, and boredom). For this new version, we observed that the scale offers a more parsimonious way to measure MA and covers a large portion of the construct without impairing the theoretical contribution from the literature (García-Santillán et al., 2018; Muñoz, Mato-Vazquez, 2007). Further observations demonstrated that ATMS-11 also has significantly higher reliability estimates (Field, 2009). Besides, the concerns that MA ought to be screened using brief, valid and reliable instruments make ATMS-11 ideal and useful (Carey et al., 2017; Suinn, Winston, 2003).

Furthermore, the results indicated that ATMS-11 is a sensitive instrument for measuring MA among males and females across cultures. We established configural, metric, and scalar invariance in samples from Mexico, Ghana, Pakistan, Iran, Romania, India, Egypt, and the overall data following the recommendations of Chen (2007). Although MI was not accessed for data from Malaysia, Nigeria, the United Arab Emirates, Thailand, and Ukraine due to data inadequacies, there is enough suggestion from our results that the usefulness of ATMS-11 cannot be underestimated.

Directions for future research

The results of the current study are presently one of the most extensive datasets on cross-national and continental MA evaluation and instrument standardisation. However, some issues limit the generalisability of our results. First, the sample pools of each country were purposively selected and may only represent a self-selected sample out of their local or national population. Second, the present study included most research sites in Asia and Africa with much less or no representations of South, North and Central America, Europe, the Caribbean, Australia, and Oceania, mainly due to the flexible recruitment method. Thus, our samples are less likely to represent their selected countries fully. Third, we could not exclude participants with a formal diagnosis of mood or learning disorders due to the nature of the design and modalities for data collection. Additional studies with larger samples from several countries will be needed in the future to address these issues. Also, longitudinal studies may provide additional understanding of developmental paths among various age categories, as this study failed to address that.

5. Conclusion

The present study was conducted to establish the validity and reliability parameters of the ATMS-24 across nations. The study also provided a modified and abbreviated version of the original scale, ATMS-11, which is useful for measuring the differences in MA between males and females from different cultures with a cross-national perspective. Using 4,338 participants from 12 countries and across four continents, the ATMS-11 is one of the current existing instruments for assessing MA with the highest multinational standardisation sample compared to several previous studies (Bai et al., 2009; Beasley et al., 2001; Carey et al., 2017; Chiu, Henry, 1990; García-Santillán et al., 2016; Muñoz, Mato, 2007). The novel 11-item scale may offer a valid, reliable, and parsimonious means to assess MA in different populations as it discriminates between male and female students across cultures. Though the data for this study has some limitations, the usefulness, validity, and reliability of ATMS-11 cannot be underestimated (see Table 5).

Table 5. Modified Scale of Anxiety Towards Mathematics (ATMS-11)

Items
1. I feel nervous when I get the math test questions.
2. I feel nervous when I think about the math test that I have to do next week.
3. I get nervous when I know that in the next course I will still have math classes.
4. I feel nervous when I think about the math test that I have next week.
5. The math exams make me nervous.
6. It makes me nervous to do mathematical operations.
7. I feel nervous when they give me a list of math exercises.
8. I feel nervous when I do a math assessment test.
9. I feel nervous when I see/hear my teacher explaining a math problem.
10. I feel nervous when I have to explain a problem in math class.
11. I feel nervous when I start doing my homework.

Note. Each item is rated on a five-point scale, from Strongly disagree (1), Disagree (2), Neutral (3), Agree (4), and Strongly agree (5).

6. Acknowledgements

Our gratitude goes to all collaborating institutions, heads, teachers, and parents for their support. We thank all participants for sharing their experiences in the study.

7. Declaration of Conflicting Interests

The authors declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

8. Funding

The authors received no financial support for this article's research, authorship, and/or publication.

References

Bai et al., 2009 – *Bai, H., Wang, L., Pan, W., Frey, M.* (2009). Measuring mathematics anxiety: Psychometric analysis of a bidimensional affective scale. *Journal of Instructional Psychology*. 36(3): 185-193.

Beasley et al., 2001 – *Beasley, T.M., Long, J.D., Natali, M.* (2001). A confirmatory factor analysis of the mathematics anxiety scale for children. *Measurement and Evaluation in Counseling and Development*. 34(1): 14-26.

Brush, 1978 – *Brush, L.* (1978). A validation study of the Mathematics Anxiety Rating Scale (MARS). *Educational and Psychological Measurement*. 38: 485-490.

Carey et al., 2017 – *Carey, E., Hill, F., Devine, A., & Szűcs, D.* (2017). The modified abbreviated math anxiety scale: A valid and reliable instrument for use with children. *Frontiers in Psychology*. 8(11). DOI: 10.3389/fpsyg.2017.00011

Chalmers, 2012 – *Chalmers, R.P.* (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*. 48(6): 1-29. DOI: 10.18637/jss.v048.i06

Chen, 2007 – *Chen, F.F.* (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*. 14(3): 464-504.

Chiu, Henry, 1990 – *Chiu, L.-H., Henry, L.L.* (1990). Development and validation of the mathematics anxiety scale for children. *Measurement and Evaluation in Counseling and Development*. 23(3): 121-127.

Dowker et al., 2016 – *Dowker, A., Sarkar, A., Looi, C.Y.* (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*. 7: e508. DOI: 10.3389/fpsyg.2016.00508

Dreger, Aiken, 1957 – *Dreger, R.M., Aiken, L.R.* (1957). The identification of number anxiety in a college population. *Journal of Educational Psychology*. 48: 344-351.

Ellwart, Konradt, 2011 – *Ellwart, T., Konradt, U.* (2011). Formative versus reflective measurement: An illustration using work-family balance. *The Journal of Psychology*. 145(5): 391-417.

Ferrando, Lorenzo-Seva, 2011 – *Ferrando, P.J., Lorenzo-Seva, U.* (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement*. *78*: 762-780. DOI: 10.1177/0013164417719308

Ferrando, Lorenzo-Seva, 2017 – *Ferrando, P.J., Lorenzo-Seva, U.* (2017). Program FACTOR at 10: origins, development and future directions. *Psicothema*. 29(2): 236-241. DOI: 10.7334/psicothema2016.304

Field, 2009 – *Field, A.P.* (2009). Discovering Statistics Using SPSS (3 ed., Vol. 8057). London, England: SAGE Publication Ltd.

García-Santillán et al., 2018 – *García-Santillán, A., García-Cabrera, R.V., Molchanova, V.S., García-Cabrera, V.* (2018). Psychometric properties of the Scale of Mato and Muñoz-Vázquez in medical undergraduate students' sample. *European Journal of Contemporary Education*. 7(2): 332-343.

García-Santillán et al., 2016 – *García-Santillán, A., Mato-Vázquez, D., Muñoz-Cantero, J.M., Rodríguez-Ortega, D.* (2016). Anxiety toward mathematics in students of the CONALEP campus Veracruz-1: A comparative study of the morning and afternoon shifts in Veracruz, México. *Revista Internacional de Investigación e Innovación en Didáctica de las Humanidades y las Ciencias*. 3: 55-75.

Hair et al., 1998 – *Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.L.* (1998). Multivariate Data Analysis. Upper Saddle River, NJ: Prentice Hall.

Hopko, 2003 – *Hopko, D.R.* (2003). Confirmatory factor analysis of the math anxiety rating scale-revised. *Educational and Psychological Measurement*. 63(2): 336-351.

Jamieson et al., 2016 – *Jamieson, J.P., Peters, B.J., Greenwood, E.J., Altose, A.J.* (2016). Reappraising stress arousal improves performance and reduces evaluation anxiety in classroom exam situations. *Social Psychological and Personality Science*. 7(6): 579-587.

Keshavarzi, Ahmadi, 2013 – *Keshavarzi, A., Ahmadi, S.* (2013). A comparison of mathematics anxiety among students by gender. *Procedia – Social and Behavioral Sciences*. 83: 542-546.

Lorenzo-Seva, Ferrando, 2019 – *Lorenzo-Seva, U., Ferrando, P.J.* (2019). Robust Promin: A method for diagonally weighted factor rotation. Technical report, URV. Tarragona, Spain. DOI: 10.24265/liberabit.2019.v25n1.08

Mahmood, Khatoon, 2011 – *Mahmood, S., Khatoon, T.* (2011). Development and validation of the mathematics anxiety scale for secondary and senior secondary school students. *British Journal of Arts and Social Sciences*. 2(2): 169-179.

Muñoz, Mato, 2007 – *Muñoz, J.M., Mato, M.D.* (2007). Elaboración and estructura factorial de un cuestionario para medir la ansiedad hacia las matemáticas en alumnos de educación secundaria obligatoria [Elaboration and factorial structure of a questionnaire to measure math anxiety in students of compulsory secondary education]. *Revista Galego-portuguesa de Psicoloxia e Eduación*. 14(11): 221-231. [in Spain]

Pajares, Urdan, 1996 – *Pajares, F., Urdan, T.* (1996). Exploratory factor analysis of The Mathematics Anxiety Scale. *Measurement and Evaluation in Counseling and Development*. 29(1): 35-47.

Polly et al., 2018 – *Polly, D., Wang, C., Martin, C., Lambert, R., Pugalee, D., Middleton, C.* (2018). The influence of mathematics professional development, school-level, and teacher-level variables on primary students' mathematics achievement. *Early Childhood Education Journal*. 46(1): 31-45.

R Core Team, 2019 – R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2019.

Rosseel, 2012 – *Rosseel, Y.* (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*. 48(2): 1-36.

Samejima, 1969 – *Samejima, F.* (1969). Estimation of latent ability using a response pattern of graded scores (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society.

Silver et al., 2021 – *Silver, A.M., Elliott, L., Libertus, M.E.* (2021). When beliefs matter most: Examining children's math achievement in the context of parental math anxiety. *Journal of Experimental Child Psychology*. 201: e104992.

Suinn, Winston, 2003 – *Suinn, R.M., Winston, E.H.* (2003). The mathematics anxiety rating scale, a brief version: Psychometric data. *Psychological Reports*. 92(1): 167-173.

Suinn, 1988 – *Suinn, R.M., Taylor, S., Edwards, R.W.* (1988). Suinn mathematics anxiety rating scale for elementary school students (MARS-E): Psychometric and normative data. *Educational and Psychological Measurement*. 48(4): 979-986.

Svetina et al., 2020 – *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.

Tetrick, Buffardi, 2006 – Tetrick, L.E., Buffardi, L.C. (2006). Measurement issues in research on the work-home interface. In F., Jones, R.J., Burke and M., Westman (Eds.), Work-life balance: A psychological perspective, (pp. 90-114). New York, NY: Psychology Press.

Timmerman, Lorenzo-Seva, 2011 – *Timmerman, M.E., Lorenzo-Seva, U.* (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*. 16: 209-220. DOI: 10.1037/a0023353

Waller, Flood, 2016 – *Waller, P.P., Flood, C.T.* (2016). Mathematics as a universal language: Transcending cultural lines. *Journal for Multicultural Education*. 10(3): 294-306.

Widaman, 2007 – *Widaman, K.F.* (2007). Common factor versus components: Principals and principles, errors and misconceptions Em. In R. Cudeck and R.C. MacCallum, Factor analysis at 100: Historical developments and future directions (pp. 177-204). London: Lawrence Erlbaum Associates.

Woodhouse, Jackson, 1977 – *Woodhouse, B., Jackson, P.H.* (1977). Lower bounds to the reliability of the total score on a test composed of nonhomogeneous items: II. A search procedure to locate the greatest lower bound. *Psychometrika*. 42: 579-591. DOI: 10.1007/bf02295980

Wu, Estabrook, 2016 – *Wu, H., Estabrook, R*. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*. 81(4): 1014-1045.