Psychometric properties, factor structure and gender and educational level invariance of the Abbreviated Math Anxiety Scale (AMAS) in Spanish children and adolescents[1]

M. Eva Martín-Puga, M. José Justicia-Galiano, M. Mar Gómez-Pérez, and Santiago Pelegrina

Department of Psychology, University of Jaén

Corresponding Author

Santiago Pelegrina. Department of Psychology. University of Jaen. Paraje Las Lagunillas s/n. 23071 Jaen. Spain

Email: spelegri@ujaen.es

Acknowledgements

This study was supported by a grant from the Spanish Ministry of Economy, Industry and Competitiveness (PSI2016-80416-P funded by FEDER).

---

[1] Accepted manuscript version. Please cite as:

Martín-Puga, M.E., Justicia-Galiano, M.J., Gómez-Pérez, M. M., & Pelegrina, S. (2022). Psychometric Properties, Factor Structure, and Gender and Educational Level Invariance of the Abbreviated Math Anxiety Scale (AMAS) in Spanish Children and Adolescents. *Assessment*, *29*(3), 425-440. https://doi.org/10.1177/1073191120980064

Supplemental material for this article is available online: https://journals.sagepub.com/doi/abs/10.1177/1073191120980064

Psychometric properties, factor structure and gender and educational level invariance of the Abbreviated Math Anxiety Scale (AMAS) in Spanish children and adolescents

Abstract

The present study aimed to investigate the factor structure and degree of measurement invariance of a Spanish adaptation of the Abbreviated Math Anxiety Scale (AMAS) in primary and secondary school students (N = 1,504 students, 46.08% males, 7-19 years of age). The results of confirmatory factor analysis corroborated the original two-factor structure, although a modified two-factor model with one item loading simultaneously on both factors was better supported. Full measurement invariance was observed across gender, and partial measurement invariance was achieved across educational levels (primary and secondary education). The AMAS showed reasonable internal consistency, test-retest reliability, and convergent validity. These results highlight the utility of the AMAS as a measure of math anxiety in primary and secondary school students whose scores can be compared by gender and educational level.

Key words: Math anxiety, factor structure, measurement invariance, psychometrics properties, children, adolescents, Abbreviated Math Anxiety Scale.

A widely accepted definition of math anxiety (MA) is “feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations” (Richardson & Suinn, 1972, p. 551). MA is a general phenomenon that shows considerable prevalence among all educational levels across different countries (Foley et al., 2017). Manifestations of MA include negative attitudes towards mathematics, and lower levels of enjoyment, motivation and confidence with mathematics, which may have negative consequences (Namkung, Peng, & Lin, 2019). MA has consistently related to a drop in math performance; moreover, it may induce avoidance of everyday mathematical situations, and in the long run biases course and career choices (Ashcraft, 2002; Ramirez, Shaw, & Maloney, 2018).

Most of the research on MA was initially conducted on college and secondary school students, with comparatively less attention being devoted to primary school children (Hill et al., 2016). Over the past decade, however, interest in MA has increased, leading to a growing number of studies aimed at determining its origin and development during school years. Initial signs of MA have been reported even in 6-9-year-old children (e.g., Ganley & MacGraw, 2016; Young, Wu, & Menon, 2012), and some research has suggested that MA may become more severe with age (Dowker, Sakar, & Looi, 2016). A meta-analysis by Hembree (1990) revealed that MA increased across the first few years of high school, before levelling off from 9 grade onwards. The results regarding MA development during primary school are, however, somewhat contradictory. Whereas some research indicates an increase in MA with age (e.g., Gierl & Bisanz, 1995), a number of studies have failed to observe differences in MA across primary school grades (Dowker, Bennet, & Smith, 2012; Ganley & MacGraw, 2016) or between primary and secondary school students (Carey, Hill, Devine, & Szűcs, 2017; Dowker et al., 2012). Sorvo et al. (2019) postulated that one reason for the inconsistent results could be that previous research assessed different aspects of MA.

Furthermore, studies have tended to focus on a limited number of grades or ages. To our knowledge, no studies have used the same instrument to assess MA across a broad range of grades, from primary to the last year of secondary school.

Gender differences in MA have been widely reported. Females are more prone to experience higher levels of MA than males in secondary and post-secondary education (e.g., Hembree, 1990; Dowker et al., 2016). However, inconsistent results have been obtained in the context of primary education. Although some studies found no gender difference in MA among primary school children (e.g. Harari, Vukovic, & Bailey, 2013), others did report a difference (e.g., Hill et al., 2016; Young et al., 2012). As with the research examining developmental trends in MA, most existing studies on gender differences tended to focus on particular educational levels and used different instruments to assess MA, where both of these factors may have contributed to the divergent results. Therefore, it would be informative to test for grade and gender differences in MA using the same questionnaire.

The Abbreviated Math Anxiety Scale

The Abbreviated Math Anxiety Scale (AMAS; Hopko, Mahadevan, Bare, & Hunt, 2003), which is one of the most widely used questionnaires to assess MA, was developed as a shorter alternative to the MARS-R (Math Anxiety Rating Scale-Revised; Plake & Parker, 1982). The AMAS is a short (nine-item) scale that assesses anxiety in math-related academic situations. The questionnaire includes two subscales that assess two dimensions: anxiety in learning situations with mathematical content (learning math anxiety, LMA), and test anxiety in situations in which mathematics is evaluated (math evaluation anxiety, MEA). The factor structure of the AMAS, which comprises two correlated latent factors, has been replicated in various adaptations (see Table 1). In addition to the original structure, several recent studies have shown that a two-factor model, in which one item loads on both factors, may provide a better fit to the data (e.g., Cipora, Szczygiel, Willmes, & Nuerk, 2015; Cipora, Willmes, Szwarc, & Nuerk, 2018; Schillinger, Vogel, Diedrich, & Grabner, 2018).

The AMAS has shown psychometric properties comparable to those of the longer original scale (MARS), as well as high convergent validity (*r* = .85, Hopko et al., 2003). These good psychometric properties, and the reduced number of items, make the AMAS a convenient and efficient instrument to measure MA, which can be employed in anxiety research involving children and adolescents. Efforts have been also directed toward adapting the original AMAS into different languages (Persian: Vahedi & Farrokhi, 2011; Italian: Primi, Busdraghi, Tomasetto, Morsanyi, & Chiesi, 2014; Polish: Cipora et al., 2015; Latin American Spanish, Brown, & Sifuentes, 2016; German: Schillinger et al., 2018 and Serbian: Sadiković, Milovanović, & Oljača, 2018). The various adaptations have evidenced good psychometric properties, reinforcing that the AMAS is a valid measure to assess MA in different populations (see Table 1). Although the AMAS was developed originally for research on adults, it has recently been adapted for use in children or adolescents, including Italian (Caviola, Primi, Chiesi, & Mammarella, 2017; Primi et al., 2014, 2020); English (Carey et al., 2017; Primi et al., 2020); Polish (Szczygiel, 2019) and Serbian versions (Milovanović & Branovački, 2020; Sadiković et al., 2018).

Table 1 about here

The majority of studies have applied the AMAS to students of a specific educational level, and it is not clear whether the results of the various adaptations of the AMAS are comparable. An important requirement when comparing MA scores across groups is measurement invariance (MI). At present, we are aware of only a single study, by Primi et al. (2014), which investigated the MI of the AMAS across different educational levels. These authors reported only weak invariance across high school and college groups of adolescents, indicating that the AMAS has the same meaning across these groups, although the scores could not be directly compared. However, the MI of the AMAS across lower educational levels remains to be determined.

With respect to the MI across gender, several studies have tested the gender invariance of the AMAS and reported contrasting results (Table 1). Vahedi and Farrokhi (2011) reported at least configural invariance across gender. Sadiković et al. (2018) found weak (or metric) gender invariance in secondary and university students, indicating that the AMAS has the same meaning across these groups. Meanwhile, Caviola et al. (2017) and Primi et al. (2014, 2020) observed strict factorial invariance in primary school, high school and university samples, such that equivalence of the AMAS would be expected when administered to males and females. Given the heterogeneity of the available data and paucity of research addressing the MI of the AMAS across gender, additional studies with children and adolescents are needed.

The present study

This study aimed to investigate the factor structure of the AMAS, and to determine whether the AMAS is invariant across educational levels and gender. Some efforts have been made towards adapting the AMAS for use in different school-age populations; however, no version that covers the full range of primary and secondary school grades has been devised. Furthermore, it is necessary to adapt the AMAS for use in Spanish-speaking children and adolescents. Spanish is a language used by more than 47 million people in Europe, and is also one of the most widely spoken around the world, with 580 million speakers. A Spanish version of the AMAS may facilitate meaningful comparison of the data of populations in different regions. Considering the good psychometric properties of the AMAS, and the absence of a validated instrument for use in a wide range of ages, we aimed to adapt the AMAS for use in Spanish primary and secondary school children.

One goal of the present study was to investigate the factor structure of the AMAS in children and adolescents. Although the original two-factor solution reported by Hopko et al., (2003) has been repeatedly validated (e.g., Caviola et al., 2017; Primi et al., 2014), several recent studies reported a model in which one item loads on both factors (e.g. Cipora et al., 2015; 2018; Schillinger el., 2018). This model has been used in adult samples, but has not yet been tested in children and adolescents. Therefore, we deemed it relevant to determine whether the recently reported model also applies to primary and secondary school children. A further aim of this study was to evaluate the MI of the AMAS across gender and educational levels. No analysis of invariance across educational levels including children and adolescents has been performed previously, and invariance across gender has been reported only for a restricted range of grades; moreover, the results have been equivocal. Hence, to assess the measurement invariance of the AMAS, we administered it to a large sample of primary (Grades 3 to 6) and secondary school children (Grades 7 to 12).

Method

This study was conducted as part of a larger investigation examining the relationship between math anxiety and math achievement, and possible mediators and moderators thereof, across different educational levels. Below, we report how we determined our sample size, all data exclusions, and all measures in the study.

Participants

Our choice of a large sample size was informed by previous studies investigating AMAS (e.g., Carey et al., 2017; Caviola et al., 2017). A subsequent power analysis performed using G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that a total of 379 participants would be needed to detect small effects, such as gender and grade,

(R^2 = .03) in a linear multiple regression analysis with two predictors, with 80% power and an alpha at .05. Thus, 100 participants per grade was set as the minimum requirement in the primary school group (four grades); this requirement also applied to the secondary school group. As there were 10 grades, the minimum sample size was 1,000 participants. A total of 1,571 children and adolescents volunteered to participate in the study, exceeding the minimum requirement. Students with special education needs, who may have difficulties in understanding the questions, were excluded from the study. Of the remaining 1,530 students, 26 left one or more AMAS item blank; these data were discarded from the analysis. The final sample consisted of 1,504 students (46.08% males), aged 7 to 19 years (*M* = 13.59; *SD* = 2.71), in Grades 3 to 12. The resulting sample size exceeded the minimum of 200 participants recommended for Monte Carlo studies of measurement invariance (Cheung & Rensvold, 2002).

Participants were split into two groups based on their educational level: primary education (Grades 3 to 6) and secondary education (Grades 7 to 12). The mean age of the primary school children (*n* = 449, 46.3% males) was 10.19 years (*SD* = 1.16; range: 7–12 years), and the mean age of the secondary school children (*n* = 1,052, 46.0% males) was 15.05 years (*SD* = 1.68; range: 11–19 years). The mean age of the male and female subsamples was 13.63 years (*SD* = 2.76) and 13.56 years (*SD* = 2.66), respectively. Three participants did not provide information on their age. The students were recruited from eight schools and high schools in two Spanish cities. The study was approved by the Ethics Committee of the University of Jaén. Written informed consent was obtained from the participants´ parents, and verbal assent was given by the children, to take part in the study.

Measures and Procedure

Abbreviated Math Anxiety Scale (AMAS). The AMAS consists of nine items divided into two subscales. Items 1, 3, 6, 7 and 9 comprise the LMA subscale and the remaining items (2, 4, 5 and 8) constitute the MEA subscale (see Appendix A). Students are asked to rate how nervous or anxious they would feel in different math situations on a 5-point Likert ranging from 1 (not nervous at all) to 5 (very nervous). The LMA score ranges from 5 to 25, and the MEA score from 4 to 20; thus, the total AMAS score ranges from 9 to 45. In all cases, higher scores indicate higher levels of MA.

The adaptation began with translation of the AMAS into Spanish (V1). The Spanish version was then back-translated into English (V2) by two native English teachers in Spanish schools. The new English version (V2) was compared to the original English version, and proved to be grammatically and semantically equivalent, although some original expressions or words had to be adapted to the less complex mathematical contents of the lower level courses. The response categories were also modified by substituting 'anxious' with 'nervous'; in the initial instructions to participants, these words were introduced as synonyms.

Math Anxiety Scale for Young Children (MASYC). The MASYC (Harari et al., 2013) consists of 12 items assessing MA that are grouped into three factors: Negative Reactions (1–3); Numerical Confidence (4–8) and Worry (9–12). Students are asked to indicate the extent to which they agree with each statement on a 4-point Likert scale ranging from 1 (not at all) to 4 (very much). The five items for the Numerical Confidence factor have the opposite valence. Scores on the MASYC range from 12 to 48, with higher scores indicating higher anxiety. In the present study, the Cronbach's alpha (α) of the MASYC, as a measure of internal consistency, was .77.

State-Trait Anxiety Inventory for Children (STAIC). The STAIC (Spielberger, Edwards, Lushene, Montuori, & Platzek, 1973) is a measure of general anxiety that comprises two 20-item subscales. Only the second scale (STAIC-T), which assesses the general level of anxiety was used in the current study. Students rate their responses on a 3-

point Likert scale ranging from 1 (almost never) to 3 (often). Scores on the STAIC-T range from 20 to 60, where higher scores indicate higher anxiety. The Cronbach's α for the STAIC-T was .86 in the current study.

Participants were tested in a classroom setting during a single session. The anxiety questionnaires described in the present study were administered first. Then, the participants completed other motivational questionnaires and finally a set of cognitive and mathematical performance tasks that are not reported in the present study. In accordance with common practice in individual differences research, the order of administration of the various measures was the same for all participants. A fixed order prevented the possibility of math performance influencing self-reported anxiety. The assessment lasted about 90 minutes for primary school children and about 60 minutes for secondary school participants. A subsample of 186 students (97 primary and 89 secondary school children) completed the scale 4-5 weeks after the first test session in order to determine the test-retest reliability of the AMAS.

Data analysis

The first stage of the data analysis involved CFA of the total AMAS scores. Three models were tested: a unidimensional model (M1), the original model (M2) for AMAS, and the recently proposed model with an item allowing loading on two factors (M3). Given that the AMAS items are ordinal, models of ordered-categorical variables might be preferable for single-group CFA (Beauducel & Herzberg, 2006). Therefore, a weighted least squares mean and variance-adjusted (WLSMV) estimator is appropriate (Flora & Curran, 2004). This is a robust estimator that does not assume that scores are normally distributed. The following indices were used for assessing model fit: the scaled Satorra–Bentler χ^2 test, the scaled comparative fit index (CFI), the scaled Tucker-Lewis Index (TLI), the scaled root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR). For simplicity, the term "scaled" will henceforth be omitted. Cutoff values close to

.95 (CFI and TLI), .06 (RMSEA) and .08 (SRMR) are indicative of a relatively good fit (Hu & Bentler, 1999). The model with the best fit was determined by comparing the models using the Satorra–Bentler scaled χ^2 difference test (Satorra & Bentler, 2001).

The second analysis stage involved determining the MI across gender and educational level considering three increasingly stringent levels of invariance. We adopted the approach proposed recently by Wu and Estabrook (2016) for setting the identification constraints of invariance models with ordered-categorical variables. In the first level, configural invariance indicates that construct dimensionality is equivalent between groups (males and females, and primary and secondary educational levels). This type of invariance was examined by determining whether the groups had the same number of factors and loading patterns. For model identification, factor variances and means were constrained to 1 and 0, respectively, in both groups. When configural invariance is observed, Wu and Estabrook recommended testing thresholds before loadings, instead of following the conventional sequence for testing measurement invariance. Thus, in the second level, threshold invariance was examined by constraining the thresholds to be equal across groups. For model identification, intercepts were fixed to 0 in the first group (reference) and freed in the second group. Finally, loading invariance was determined by constraining the loadings to be equal across the groups. Factor variance was set to 1 in the reference group and freed in the second group, for identification purposes. Theta parametrization was used.

When loading or threshold invariance was not supported, partial invariance models were tested. In these models, some parameters (e.g., loadings, thresholds) were constrained to be equal, while others were allowed to vary across groups (Putnich & Bornstein, 2016). The parameters to be released were identified based on the modification indices and Lagrange multiplier test provided by the lavTestScore function in semTools. This statistic is informative regarding the effect of releasing an equality constraint between groups.

Parameters were released one at a time. Partial measurement invariance was considered acceptable if the proportion of noninvariant parameters did not represent more than the 20% of the total number of parameters (loadings and thresholds) (Dimitrov, 2010). In addition, more than half of the items comprising a given factor should be invariant for valid group comparisons (Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000).

Chi-square difference testing has been used to compare the fit of measurement invariance models. However, given that the likelihood ratio test χ^2 is sensitive to sample size, some authors (e.g., Chen, 2007; Cheung & Rensvold, 2002) have advocated the use of differences in approximate fit indices (AFIs), like CFI and RMSEA values, to test models with continuous data. Unfortunately, the utility of these indices has not been extensively studied in models with ordered-categorical indicators. Sass, Schmitt and Marsh (2014) found that the aforementioned cut-off criteria for change in AFIs performed poorly with the WLSMV estimator, especially in models with some degree of misspecification. In the context of invariance in ordered-categorical CFA models, $\Delta\chi^2$ estimated using WLSMV has shown high power and sensitivity to detect metric invariance, although may be associated with minor inflation of the Type I error rate (Koziol & Bovaird, 2018; Sass et al., 2014). Given that some authors have recommended that relying exclusively on ΔAFIs be avoided (Han, Colarelli & Weed, 2019; Liu et al., 2017; Sass et al., 2014), we opted to apply a conservative criterion and continue comparing models in the event of a significant $\Delta\chi^2$. The R code used for conducting the CFAs and assessing measurement invariance is provided in Appendix A in the Supplementary Material.

The psychometric properties of the subscales and the overall scale were analyzed. Internal consistency was estimated through McDonald's omega coefficient. The omega coefficient offers some advantages over the classical Cronbach alpha, in that it does not assume that each indicator contributes equally to the factor loadings (tau-equivalence), and the error variances of the indicators are uncorrelated. The test-retest reliability of the AMAS and its subscales was also calculated.

Discriminant and convergent validity were estimated based on the correlations between AMAS scores, other math anxiety measures (the Worry and Negative Reactions MASYC scales), and a general anxiety measure (STAIC-T), as well as attitude-related items (Numerical Confidence scale from MASYC). We compared the correlation coefficients using the Pearson and Filon's *z* test, implemented in the cocor package (Diedenhofen & Musch, 2015). In a complementary approach to investigate discriminant validity, we determined the heterotrait-monotrait ratio of correlations (HTMT, Henseler, Ringle, & Sarstedt, 2015) between the different measures considered.

Normative data stratified by gender and grade were obtained, and the possible effects of grade and gender on the different scales were determined. Finally, a multiple linear regression analysis was conducted to determine the effects of gender, grade and educational level, and the first order interactions involving educational level, on the LMA and MEA scores. Gender was coded as 0 for female and 1 male, and educational level was coded as 0 for primary and 1 for secondary education.

All analyses were conducted with R (version 3.6.2; R Core Team, 2019). CFA and measurement invariance models were fit with lavaan (version 0.6-7; Rosseel, 2012). Parameters to be released in measurement invariance analyses, McDonald's omega coefficients and the HTMT ratios of the correlations were obtained with seemTools (version 0.5-3; Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2020).

Results

Confirmatory Factor Analysis

Three models were evaluated: Model 1 (M1) included only one factor. Model 2 (M2), included two correlated factors (Learning Math Anxiety and Math Evaluation Anxiety), as proposed by the original AMAS model (Hopko et al., 2003). Model 3 (M3) involved two correlated factors with Item 5 loading on both factors, as reported by Cipora et al. (2015), Cipora et al. (2018) and Schillinger et al. (2018).

Table 2 presents the fit indexes for all the models. χ^2 was significant and the χ^2/df ratio exceeded the cutoff criteria for an acceptable fit for all the models. It should be noted that the use of these indices has been discouraged, while alternative fit indices, as reported in Table 2, have been recommended (e.g., Brown, 2015). The model fit analysis according to these indices showed that the one-factor model (M1) fitted the data poorly. The two-factor model reflecting the original structure of AMAS (M2) fitted the data reasonably well. The model (M3), in which Item 5 loaded on the two factors, also showed acceptable fit. The model comparison indicated that the model with two factors (M2) fitted significantly better than the one-factor model (M1) (SBχ^2difference test (1) = 184.63, $p < .001$). Furthermore, M3, with Item 5 cross-loading on both factors, showed a better fit than the two-factor model (M2) (SBχ^2difference test (1) = 40.62, $p < .001$).

Table 2 about here

Table 3 presents the factor loadings for each model, where the values ranged from .31 to .83. The correlations between both factors were .77 and .73 for M2 and M3, respectively. Even though model M3 showed superior fit, it had lower theoretical parsimony than the originally hypothesized model (M2) due to cross-loading of Item 5. In addition, inspection of the standardized loadings of Item 5 in M3 revealed that they were weak and similar in magnitude (0.32 and 0.34). The weak loading of Item 5 in model M3, as also reported in other studies that have explored this model (Cipora et al., 2015, Schillinger et al., 2018), make this model difficult to interpret. In contrast, Item 5 had an acceptable loading in model

M2. Therefore, considering the model fit, magnitude of cross-loadings and model theoretical parsimony, the original two-factor model (M2) was preferred for further analyses of measurement invariance.

Table 3 about here

Measurement invariance across gender and educational level

The MI for the original model (M2) was examined across gender groups (female and male) using the WLSMV estimator. As shown in Table 4 the configural model fit the data acceptably. Threshold invariance was supported, as evidenced by a non-significant $\Delta\chi^2$ (p = .40) and an acceptable model fit when these parameters were constrained. Scalar invariance with loadings and thresholds constrained to be equal across gender was also obtained, given that $\Delta\chi^2$ was non-significant (p = .54) and the model fit the data well. Therefore, the AMAS can be used to obtain MA scores that are amenable to comparison between children and adolescents, of both genders.

Table 4 about here

With regard to invariance across educational level, the fit of the configural model was reasonably good (see Table 4), indicating a common structure invariant for all groups considered. Threshold invariance was not supported because the fit of the model decreased when thresholds were constrained to be equal across groups ($\Delta\chi^2$ = 36.93, p = .005). Inspection of the test statistics showed that the first threshold constraint on Item 2 (thinking about an upcoming math test 1 day before) should be freed. As a result, the first threshold associated with Item 2 was higher in the secondary (-0.68) than in the primary school group

(-1.0). Imposing this partial invariance constraint did not result in a deterioration in model fit when compared with the configural invariance model ($\Delta\chi^2 = 23.27$, $p = .14$) and the model fit the data well. In the next step, after constraining the loadings to be equal between the two groups, scalar invariance (thresholds and loadings) was not supported ($\Delta\chi^2 = 25.21$, $p = .001$). Based on inspection of the test statistics, the loading for Item 1 (having to use the tables in the back of a math book) was allowed to freely vary across groups. After releasing this equality constraint, partial scalar invariance was achieved, as evidenced by a nonsignificant change in fit ($\Delta\chi^2 = 8.09$, $p = .231$) and a good model fit. In total, two of the 45 (4.4%) possible parameters were released.

Reliability and validity

Reliability was estimated using McDonald's omega coefficient. The omega value was .86 for the AMAS scale (.81 for primary and .88 for secondary school children), .75 for the LMA subscale (.61 for primary and .80 for secondary school children), and .82 for the MEA subscale (.78 for primary and .84 for secondary school children). Alpha ordinal and alpha values are given in Table 1 to allow comparison with the same statistics reported in previously published studies. It should be noted that classical alpha may underestimate score reliability under violations of tau-equivalence (Dunn, Baguley and Brunsden, 2013).

The 4/5-week test-retest reliability was .74 for the AMAS scale (.71 for primary and .76 for secondary school children), .66 for the LMA subscale (.66 for primary and .65 for secondary school children), and .71 for and MEA subscale (.68 for primary and .73 for secondary school children).

In order to investigate the convergent and divergent validity of the AMAS, correlation coefficients among the two subscales of the AMAS, two math anxiety measures (Negative Reactions and Worry from MASYC), a measure of general anxiety (STAIC-T) and an additional measure related to math attitudes (Numerical Confidence from MASYC, with reverse scoring of items) were computed. As shown in Table 5, moderate correlations were found between the LMA and MEA scores and math anxiety, as indexed by the MASYC Worry and Negative Reactions subscales. The correlation coefficients between LMA and the two MA measures from MASYC were larger than the coefficient between the LMA and STAIC-T scores (both *z*s > 2.46, both *p*s < 0.014). The correlation coefficient between MEA and Worry was also larger than that between MEA and STAIC-T (z = 5.08, p < .001) but, somewhat surprisingly, the correlation between MEA and Negative Reactions did not differ significantly from that between MEA and STAIC-T (z = -1.39, p =.16). Finally, the correlations between the LMA and MEA scores and Numerical Confidence, assessed using the MASYC subscale, were weaker than those observed between any of the MA measures (all *z*s > 6.30; all *p*s < .001), providing additional support for discriminant validity. The complementary heterotrait-monotrait analysis showed that the HTMT ratios of correlations between the AMAS subscales and the two math anxiety measures from MASYC ranged from .74 to .82, with higher ratios for Worry than for Negative Reactions. Notably, all of these HTMT ratios were larger than those observed between the two AMAS measures and general anxiety (STAIC-T). The latter ratios were in turn larger than the corresponding HTMT ratios between the AMAS subscales and Numerical Confidence scale. Together, these results suggest that the AMAS provides a valid measurement of math anxiety.

Table 5 about here

Normative data and gender and grade differences

Table 6 lists the descriptive statistics for the nine AMAS items. Percentile values and descriptive statistics for the two subscales and the total scores, stratified by grade and gender, are presented in Appendix B.

Table 6 about here

Finally, two multiple linear regression analyses were conducted to determine the effects of gender, grade and educational level on the scores of each AMAS subscale. As can be seen from Table 7, gender had a similar effect on each measure, where female students consistently reported higher levels of math anxiety than male students. The relationship between grade and each measure of MA depended on the educational level. MEA scores increased steadily with grade throughout primary and secondary school. In contrast, the effect of grade on LMA scores differed by educational level, as evidenced by the significant interaction. Separate analyses for each educational level revealed that LMA scores increased with grade in the context of secondary education ($\beta = 0.18$, $p < .001$), whereas for primary education, LMA scores exhibited a slight decline ($\beta = -0.09$, $p = .052$).

Table 7 about here

Discussion

The objective of this study was to analyze the factor structure of a Spanish version of the AMAS, by applying it to 1504 children and adolescents in Grades 3 through 12, and to test the invariance of the scale across age and gender. Overall, the data supported a two-factor structure. Additionally, the results provided evidence of scalar MI across gender and partial scalar invariance across educational level.

Factor Structure

The results of the CFA showed poor fit of the one-factor model (M1), thereby casting doubt on the unidimensionality of AMAS. For this reason, we refrained from computing a total score for the instrument. In contrast, our findings revealed an adequate fit of the two-factor model (M2), which included a factor for LMA and another for MEA. This model is consistent with the original two-factor structure obtained by Hopko et al. (2003) that has been confirmed in various AMAS adaptations (Brown & Sifuentes, 2016; Carey et al., 2017; Caviola et al., 2017; Cipora et al., 2015; 2018; Milovanović & Branovački, 2020; Primi et al., 2014; 2020; Sadiković et al., 2018, Schillinger et al., 2018; Szczygiel, 2019; Vahedi & Farrokhi, 2011). All of the items showed strong factor loadings (from .64 to .81) except for Item 1, which loaded weakly on the LMA factor. Existing studies that have examined the factor structure of the AMAS reported somewhat inconsistent loadings for this item, with values ranging from .09 to .56, a point we discuss in more detail below.

The model (M3) in which Item 5 was allowed to load on two factors showed an even better fit than the original two-factor model (M2), which is also in line with results recently reported for adult samples by Cipora et al. (2015, 2018) and Schillinger et al. (2018). Nonetheless, a limitation of this model is the weak cross-loading of Item 5; although similar to the values close to .4 reported by Cipora et al. (2015, 2018) and Schillinger et al., (2018), it creates some ambiguity in the model interpretation. Item 5 refers to a situation in which difficult homework is assigned for the next scheduled class; this could elicit evaluation anxiety as the schoolwork will be reviewed by a teacher, but may also elicit learning anxiety given the task difficulty. Therefore, because specifying the cross-loading makes the model M3 less parsimonious and more difficult to interpret, the original model M2 was preferred for testing measurement invariance.

Measurement invariance across gender

Further, we examined the MI for model M2 in the factor structure of the AMAS across gender and educational levels. For gender groups, scalar (loadings and thresholds) invariance was supported. Caviola et al. (2017) and Primi et al. (2014, 2020) observed a similar level of invariance for gender in primary and secondary students, while Sadiković et al. (2018) reported only weak invariance (i.e., loadings) across gender. Overall, the present results indicate that scores on the latent factors of the AMAS can be directly compared irrespective of subject gender.

We observed gender differences in MA in both the primary and secondary education samples, with females reporting higher levels of anxiety than males. This result corroborates previous research showing heightened MA in women, and is in line with most previous AMAS adaptations studies (Carey et al., 2017; Caviola et al., 2017; Cipora et al., 2015, 2018; Primi et al., 2014; Schillinger et al., 2018; and Szczygiel, 2019). Several mechanisms operating in a similar manner among different age groups may underlie gender differences in MA, including gender stereotypes (Passolunghi, Ferreira, & Tomasetto, 2014) and spatial processing abilities (Maloney, Ansari, & Fugelsang, 2011). However, it is interesting to note that analogous gender differences were not found in some of the AMAS adaptations (Milovanović & Branovački, 2020; Primi et al., 2020, Sadiković et al., 2018, and Vahedi & Farrokhi, 2011). This variation across studies using the same instrument suggests a contribution of social and cultural factors to MA that may lead to cross-national differences therein. An example of such factor is gender equality. Stoet, Bailey, Moore and Geary (2016) showed that, in countries with higher levels of gender equality, larger gender differences in MA exist.

Measurement invariance across educational levels

Configural invariance was achieved across educational levels, indicating that the items load on the same factors in both groups. Full threshold invariance was not supported, but partial threshold invariance was achieved when the first threshold for Item 2 (thinking about an upcoming math test 1 day before) was allowed to vary across groups. This suggests that children in both groups with comparable levels of math evaluation anxiety may respond differently to Item 2. Primary school children with a true score of -1.0 for math anxiety are more likely to endorse the second category of response for this item, whereas secondary school children with the same score for anxiety are more likely to endorse the first response option. We believe that this small inequality does not have serious implications for the interpretation of the math anxiety scores.

Full loading invariance was not observed, although partial loading invariance was obtained after releasing the load constraint on Item 1. Thus, the relationship between the item and its factor differed across educational levels. Previous research has shown considerable variability in the loadings reported for this item. Whereas studies conducted with samples of adults found moderate loadings for Item 1 (Hopko et. al, 2003; Cipora et al., 2015; and Schillinger et al., 2018), studies using samples of children or adolescents reported weak or even very low loadings (Caviola et al., 2017; Sadiković et al., 2018). A possible reason for the considerable variability and low-to-moderate loading values might be that the content of this item pertains to the use of tables in the back of a math book, which may be unfamiliar to some groups of students. Indeed, the wording of this item has been modified in AMAS versions adapted for use in young children (Carey et al., 2017, Primi et al., 2020), and three studies have found that the item with the alternative wording loaded strongly on its factor (Carey et al., 2017; Milovanović & Branovački, 2020; and Szczygiel, 2019; although see Primi et al., 2020).

Because only one loading and one threshold were found to be noninvariant, the proportion of parameters released was below the criterion specified for acceptable partial scalar invariance (Dimitrov, 2010). Therefore, we may assume that the constructs are calibrated similarly across groups, which make it possible to compare the group means. This suggests that the AMAS might be useful in both primary and secondary school populations, and that their scores could be directly compared.

The analysis of possible changes in MA across grade in primary and secondary school children showed a different trajectory of the scores for each scale. Whereas MEA scores increased as a function of grade in both stages, LMA scores showed a slight decline with grade in primary school children, which changed to an upward trend with grade in the secondary school population. This finding points to the importance of considering the multifaceted nature of math anxiety, and may in part explain the different patterns of change in MA previously reported (see Dowker et al., 2016; Ramirez et al., 2018). The negative effect of evaluation seems to accumulate at an early stage, suggesting that testing is a source of anxiety from the beginning of schooling. This accords with the trajectory reported most commonly in cross-sectional studies (Dowker et al., 2016; Hembree, 1990). In contrast, the negative effect of learning experiences begins to accrue later, probably associated with the transition to secondary education, which is accompanied by a decline in performance, enjoyment and self-efficacy with respect to math (e.g., Evans, Borriello, & Field, 2018). This may help to explain why a number of studies have failed to report clear differences in MA across primary school years (Ramirez et al., 2018).

Reliability and validity

The internal consistency of the AMAS subscales was generally good for the complete sample, as well as for the primary and secondary school children groups. The reliability estimates are high and comparable to those reported for other language versions of the AMAS (Brown & Sifuentes, 2016; Carey et al., 2017; Caviola et al., 2017; Cipora et al., 2015, 2018; Primi et al., 2014, 2020; Milovanović & Branovački, 2020; Sadiković et al., 2018, Schillinger et al., 2018; Szczygiel, 2019; Vahedi & Farrokhi, 2011). Only the internal consistency of the LMA subscale was found to be moderate for the primary school group ($\omega = .61$), which may be a limitation of the scale. In general, LMA shows slightly lower reliability than MEA across different studies, and this trend seems to be more pronounced for younger children in certain contexts. Caviola et al. (2017) and Szczygiel (2019) also observed moderate reliabilities ($\alpha = .64$ and .59, respectively) for this scale in samples of Italian primary and Polish school children, respectively. However, this bias towards lower reliability was not present in the studies by Carey et al. (2017), and Milovanović and Branovački (2020), of English and Serbian children, respectively. Finally, the test-retest reliability of the Spanish version of the AMAS is similar to that obtained previously in adult samples (Cipora et al., 2015; Schillinger et al., 2018), suggesting that the constructs are relatively stable.

Regarding validity, the LMA and MEA scores were related to other math and general anxiety measures, suggesting that variance attributable to anxiety is shared by all of the scales. At the same time, the strength of correlations with the rest of the anxiety measures was moderate, reflecting a degree of independence among the underlying constructs. The fact that three of the four correlations with other MA measures were statistically larger than those observed with general anxiety indicates that the form of anxiety assessed by the instrument is more related to math than to general anxiety (see also Ashcraft & Ridley, 2005; and Hembree, 1990). The HTMT ratio more clearly showed gradual divergence between the AMAS subscales and the other math and general anxiety measures, suggesting some degree of divergent validity. This result is consistent with findings from a small number of studies that compared the relationship between AMAS scores with at least two measures of math and general anxiety (Sadiković et al., 2018; Schillinger et al., 2018; and Szczygiel, 2019). Finally, the scores for both subscales showed a slightly stronger relationship with Worry than with Negative Reactions scores, the former reflecting concerns related to poor performance and the latter involving the physiological arousal response. The worry component of anxiety is the component most strongly related to cognitive (Moran, 2016) and math performance (Hembree, 1990).

Strengths, limitations and future directions

Our analyses revealed some weaknesses in the instrument. After almost two decades since the publication of the AMAS by Hopko et al., (2003), Item 1 has become outdated. Today, the use of computers and mobiles devices has rendered table consultation unnecessary, and they are often not included in textbooks. Fortunately, an alternative item has been proposed by researchers in some countries. Carey et al. (2017) developed an adaptation for British primary school children that has subsequently been translated into other languages (Milovanović & Branovački, 2020; Szczygiel, 2019). It would be feasible to test this item with older children, or using two versions of Item 1 for different stages. However, because the latter option would make the AMAS less comparable, we recommend employing a unique and valid item for different ages. Relatedly, it is possible that the substitution of Item 1 with an item that better represents the factor could improve the reliability of the LMA scale, which was found to be modest for the primary school group. Otherwise, it may be necessary to investigate the reason for the comparatively lower reliability of LMA in primary school children. A second limitation of relatively minor importance is the fact that Item 5 may load simultaneously in both scales, resulting in a model with cross-loadings that may fit better the data than the model originally proposed (see also Cipora et al., 2015, 2018; Schillinger et al., 2018), meaning that the item would not unequivocally represent its latent factor. We suggest that this item be retained in its current form such that scores are obtained for each scale (see also Cipora et al., 2015, 2018; and Schillinger et al., 2018). However, modification of the item wording, in order to link it exclusively to the evaluation factor, should be investigated. We also encourage future investigations to assess cross-cultural measurement invariance, to determine the equivalence of the instrument across countries and cultural groups. Primi et al. (2020) addressed this issue, reporting equivalence in Italian and English between groups of young Italian and Northern-Irish children. This approach could be applied to other languages.

One of the strengths of the present work is that it addressed a gap in the literature; to our knowledge, this is the first study to test the measurement invariance of the AMAS in primary and secondary school children, and importantly, to show that AMAS is partially invariant across these educational stages. Another strength of the present research is the large sample size, encompassing a wide range of grades (from the third grade of primary school to the last grade of secondary school), in which mathematics is typically a compulsory subject. This made it possible to observe two different trajectories for learning and evaluation anxiety, and allowed us to obtain normative data stratified by gender and grade; such data could aid math anxiety evaluations across a wide range of grades. Also, we corroborated the validity of the instrument by using various anxiety measures. Finally, instead of considering the data as continuous, we estimated the models using the WLSMV estimator, which may be more appropriate for ordinal data such as the Likert-type items of the AMAS, especially when evaluating the overall model fit (Koziol & Bovaird, 2018). The findings outlined above reinforce the strengths that the AMAS has already demonstrated as a measure of math anxiety, by showing that it is a brief, valid and reliable instrument adequate for use instead of other, longer math anxiety questionnaires.

There are a number of limitations to the present study. First, we used convenience sampling, and all participant were from the same region in the south of Spain. This may limit the generalizability of the results. Further, the present AMAS version may not be valid in linguistic contexts other than the Spanish spoken in Spain. Second, given that we relied on self-report questionnaires, our results may be prone to bias (e.g., due to social desirability). A multitrait-multimethod methodology could provide additional relevant information regarding validity. Third, we used a cross-sectional approach, which enabled us to include a wide range of grades. A longitudinal design, even with a restricted number of grades, can clearly reveal developmental trends and allow the longitudinal invariance of the AMAS to be tested. Fourth, we administered the questionnaires in a fixed order. Although this conventional practice can avoid undesirable carry-over effects, it may induce other order effects.

Conclusions

In summary, in the present study, the AMAS was applied to Spanish children and adolescents. Although the AMAS has been translated into other languages, previous adaptations did not cover an age range as extensive as that in the present study (7-19 years). The present analyses indicate that the AMAS has reasonable internal consistency and test-retest reliability, as well as acceptable convergent validity. Our results suggest a two-factor model including LMA and MEA. This factor structure was invariant across gender and partially invariant across educational school levels. Therefore, AMAS scores may be compared across gender and across educational levels.

References

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181-185. https://doi.org/10.1111/1467- 8721.00196

Ashcraft, M. H., & Ridley, K. S. (2005). Math anxiety and its cognitive consequences: A tutorial review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 315–327). Psychology Press.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of Maximum Likelihood versus Means and Variance Adjusted Weighted Least Squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd edition). The Guilford Press.

Brown, J. L., & Sifuentes, L. M. (2016). Validation study of the Abbreviated Math Anxiety Scale: Spanish adaptation. *Journal of Curriculum and Teaching, 5*(2), 76-82. https://doi.org/10.5430/jct.v5n2p76

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. https://doi.org/10.3389/fpsyg.2017.00011

Caviola, S., Primi, C., Chiesi, F., & Mammarella, I.C. (2017). Psychometric properties of the Abbreviated Math Anxiety Scale (AMAS) in Italian primary school children. *Learning and Individual Differences, 55*, 174-182. https://doi.org/10.1016/j.lindif.2017.03.006

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 464-504. https://doi.org/10.1080/10705510701301834

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

Cipora, K., Szczygiel, M., Willmes, K., & Nuerk, H. C. (2015). Math anxiety assessment with the abbreviated math anxiety scale: applicability and usefulness: insights from the polish adaptation. *Frontiers in Psychology,* 6:1833. https://doi.org/10.3389/fpsyg.2015.01833

Cipora, K., Willmes, K., Szwarc, A., & Nuerk, H. C. (2018). Norms and validation of the online and paper-and-pencil versions of the Abbreviated Math Anxiety Scale (AMAS) for polish adolescents and adults. *Journal of Numerical Cognition, 3*(3), 667-693 https://doi.org/10.5964/jnc.v3i3.121

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*, *10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and evaluation in counseling and development*, *43*(2), 121–149. https://doi.org/10.1177/0748175610373459

Dowker, A., Bennett, K., & Smith, L. (2012). Attitudes to mathematics in primary school children. *Child Development Research, 2012,* 1-8. https://doi.org/10.1155/2012/124939

Dowker. A., Sarkar, A., & Looi, C. Y. (2016). Mathematics Anxiety: What have we learned in 60 Years? *Frontiers in Psychology,* 7:508. https://doi.org/10.3389/fpsyg.2016.00508

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Evans, D., Borriello, G. A., & Field, A. P. (2018). A review of the academic and psychological impact of the transition to secondary education. *Frontiers in psychology, 9*, 1482. https://doi.org/10.3389/fpsyg.2018.01482

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The math anxiety-performance link: A global phenomenon. *Current Directions in Psychological Science, 26*(1), 52-58. https://doi.org/10.1177/0963721416672463

Ganley, C.M., & McGraw, A.L. (2016). The Development and validation of a revised version of the Math Anxiety Scale for young children. *Frontiers in Psychology,* 7:1181. https://doi.org/10.3389/fpsyg.2016.01181

Gierl, M. J., & Bisanz, J. (1995). Anxieties and attitudes related to mathematics in grades 3 and 6. *The Journal of Experimental Education, 63*(2), 139-158. https://doi.org/10.1080/00220973.1995.9943818

Han, K., Colarelli, S. M., & Weed, N. C. (2019). Methodological and statistical advances in the consideration of cultural diversity in assessment: A critical review of group classification and measurement invariance testing. *Psychological Assessment*, *31*(12), 1481–1496. https://doi.org/10.1037/pas0000731

Harari, R. R., Vukovic, R. K., & Bailey, S. P. (2013). Mathematics anxiety in young children: An exploratory study. *Journal of Experimental Education, 81*(4), 538-555. https://doi.org/10.1080/00220973.2012.727888

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for research in mathematics education, 21*(1), 33-46. https://doi.org/10.2307/749455

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2016). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. *Learning and Individual Differences, 48*, 45-53. https://doi.org/10.1016/j.lindif.2016.02.006

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale (AMAS) construction, validity, and reliability. *Assessment, 10*(2), 178-182. https://doi.org/10.1177/1073191103010002008

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2020). semTools: Useful tools for structural equation modeling. R package version 0.5-3. Retrieved from https://CRAN.R-project.org/package=semTools

Koziol, N. A., & Bovaird, J. A. (2018). The impact of model parameterization and estimation methods on tests of measurement invariance with ordered polytomous data. *Educational and Psychological Measurement*, *78*(2), 272–296. https://doi.org/10.1177/0013164416683754

Liu, Y., Millsap, R. E., West, S. G., Tein, J.-Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, *22*(3), 486–506. https://doi.org/10.1037/met0000075

Maloney, E. A., Ansari, D., & Fugelsang, J. A. (2011). Rapid communication: The effect of mathematics anxiety on the processing of numerical magnitude. *Quarterly Journal of Experimental Psychology, 64*, 10-16. https://doi.org/10.1080%2F17470218.2010.533278

Milovanović, I., & Branovački, B. (2020). Adaptation and psychometric evaluation of Modified Abbreviated Math Anxiety Scale for children in Serbia. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-020-10066-w

Moran, T. P. (2016). Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychological Bulletin*, *142*(8), 831–864. https://doi.org/10.1037/bul0000051

Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. *Review of Educational Research, 89*(3), 459-496. https://doi.org/10.3102/0034654319843494

Passolunghi, M. C., Ferreira, T. I. R., & Tomasetto, C. (2014). Math–gender stereotypes and math-related beliefs in childhood and early adolescence. *Learning and Individual Differences, 34*, 70-76. https://doi.org/10.1016/j.lindif.2014.05.005

Plake, B., & Parker, C. (1982). The development and validation of a revised version of the Mathematics Anxiety Rating Scale. *Educational and Psychological Measurement, 42*(2), 551-557. https://doi.org/10.1177/001316448204200218

Primi, C., Busdraghi, C., Tomasetto, C., Morsanyi, K., & Chiesi, F. (2014). Measuring math anxiety in Italian college and high school students: Validity, reliability and gender invariance of the Abbreviated Math Anxiety Scale (AMAS). *Learning and Individual Differences, 34*, 51-56. https://doi.org/10.1016/j.lindif.2014.05.012

Primi, C., Donati, M. A., Izzo, V. A., Guardabassi, V., O'Connor, P. A., Tomasetto, C., & Morsanyi, K. (2020). The early elementary school abbreviated math anxiety scale (the EES-AMAS): a new adapted version of the AMAS to measure math anxiety in young children. *Frontiers in Psychology*, 11:1014. https://doi.org/10.3389/fpsyg.2020.01014

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist, 53*(3), 145-164. https://doi.org/10.1080/00461520.2018.1447384

Rhemtulla, M., Brosseau-Liard, P.É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315

Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology, 19*(6), 551–554. https://doi.org/10.1037/h0033456

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2). https://doi.org/10.18637/jss.v048.i02

Sadiković, S; Milovanović, I., & Oljača, M. (2018). Another psychometric proof of the abbreviated math anxiety scale usefulness: IRT analysis. *Primenjena Psihologija*, *11*(3), 301-323. https://doi.org/10.19090/pp.2018.3.301-323

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(2), 167–180. https://doi.org/10.1080/10705511.2014.882658

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507-514. https://doi.org/10.1007/BF02296192

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, *31*(12), 1428–1441. https://doi.org/10.1037/pas0000623

Schillinger, F. L., Vogel, S. E., Diedrich, J., & Grabner, R. H. (2018). Math anxiety, intelligence, and performance in mathematics: Insights from the German adaptation of the Abbreviated Math Anxiety Scale (AMAS-G). *Learning and Individual Differences, 61*, 109-119. https://doi.org/10.1016/j.lindif.2017.11.014

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., ... & Aro, M. (2019). Development of math anxiety and its longitudinal relationships with arithmetic achievement among primary school children. *Learning and Individual Differences, 69*, 173-181. https://doi.org/10.1016/j.lindif.2018.12.005

Spielberger, C. D., Edwards, C. D., Lushene, R. E., Montuori, J., & Platzek, D. (1973). *STAIC preliminary manual*. Palo Alto, Calif.: Consulting Psychologists Press, t973.

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

Stoet, G., Bailey, D. H., Moore, A. M., & Geary, D. C. (2016). Countries with higher levels of gender equality show larger national sex differences in mathematics anxiety and relatively lower parental mathematics valuation for girls. *PloS one, 11*(4). https://doi.org/10.1371/journal.pone.0153857

Szczygieł, M. (2019). How to measure math anxiety in young children? Psychometric properties of the modified Abbreviated Math Anxiety Scale for elementary children (mAMAS-E). *Polish Psychological Bulletin*, *50*(4), 303-315. https://doi.org/10.24425/ppb.2019.131003

Vahedi, S. & Farrokhi, F. (2011). A confirmatory factor analysis of the structure of abbreviated math anxiety scale. *Iranian Journal of Psychiatry, 6*(2), 47-53.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-70. https://doi.org/10.1177/109442810031002

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. https://doi.org/10.1007/s11336-016-9506-0

Young, C. B., Wu, S. S., & Menon, V. (2012). The neurodevelopmental basis of math anxiety. *Psychological Science, 23*(5), 492-501. https://doi.org/10.1177/0956797611429134

Table 1

Summary of study characteristics for prior adaptations of the AMAS

Authors, year	Sample	Age / Grade	Country	Cronbach's α (Ordinal α)	Estimator	Model	Corr.	Measurement invariance
Hopko et al., 2003	• EFA subsample: 206 undergraduates from 1239 (41.16% males)	*M* = 19.6 (3.0) (from the entire sample)	EE.UU.	AMAS: .90 LMA: .85 MEA: .88		EFA: Two-factor	.62	
	• CFA subsample: 218 undergraduate students	*M* = 19.6 (3.0) (from the entire sample)		AMAS: .83 LMA: .74 MEA: .81		Two-factor		
Vahedi & Farrokhi, 2011	298 undergraduate students (44.63% males)		Iran	AMAS: .82 LMA: .75 MEA: .79		Two-factor	.67	Configural invariance across gender
Primi et al., 2014	• High school: 215 students (52%males)	(14-19 years) *M* = 15.9 (.79)	Italy	AMAS: .86 LMA: .81 MEA: .80	MLM	Two-factor	.54	Strict invariance across gender[b] Weak invariance across educational levels
	• College sample: 249 students (43% males)	*M* = 20.9 (3.7)		AMAS: .85 LMA: .80 MEA: .83		Two-factor	.55	
Cipora et al., 2015	857 undergraduate students (18.67% males)	(18-49 years) *M* = 21.6 (4.1)	Poland	AMAS: .85 (.88) LMA: .78 (.84) MEA: .84 (.87)	ADF	M1: Two-factor M2: Two-factor with Item 5 in common	.80 .63	

Brown & Sifuentes, 2016	• EFA sample: 404 undergrad. (32.9% males)		Mexico	AMAS: .92 LMA: .88 MEA: .87		EFA: Two-factor	.72	
	• CFA sample 400 undergrad. (34.75% males)			AMAS: .90 LMA: .88 MEA: .85		Two-factor		
Carey et al., 2017[a]	• 824 primary school children (51% males)	Grade 4 (8-9 years); *M* = 109.4 months (3.7)	UK	AMAS: .85 (.89) LMA: .74 (.81) MEA: .78 (.83)	WLSMV	Two-factor	.90	
	• 922 secondary school children (50%males)	Grades 7-8 (11-13 years); *M* = 148.1 months (4.0)		AMAS: .86 (.89) LMA: .80 (.85) MEA: .81 (.84)		Two-factor	.74	
Caviola et al., 2017	1013 primary school children (51% males)	(8-11 years); Gr. 3: *M* = 8.69 (0.7) Gr. 4: *M* = 9.49 (0.6) Gr. 5: *M* = 10.2 (0.8)	Italy	AMAS: .77 LMA: .64 MEA: .74	MLM	Two-factor	.64	Strict invariance across gender[b]
Cipora et al., 2018	615 undergraduate students and employees (32.03% males)	(17-50 years) *M* = 22 (3.9)	Poland	AMAS: .84 (.88) LMA: .72 (.80) MEA: .85 (.88)	ADF	M1: Two-factor M2: Two-factor with Item 5 in common	.55 .62	
Schillinger et al., 2018	341 undergraduate students (35.19% males)	(18-35 years) *M* = 22.06 (3.40)	Austria	AMAS: .89 LMA: .87 MEA: .86	MLM	M1: Two-factor M2: Two-factor with Item 5 in common	.59 .55	
Sadiković et al., 2018	514 high school children (45.3 males)	(15-19 years) *M* = 16.7	Serbia	Gener.: .85 (.84) LMA: .80 (.62) MEA: .76 (.60)	ML	Bifactor: One general and two specific factors		Weak invariance across gender

Szczygiel, 2019[a]	• Cross-sectional: 241 children (44.40% males) • Longitudinal: 369 children (44.44% males)	(6.1-11.2 years) Grade 1: *M* = 7.3 Grade 2: *M* = 8.2 Grade 3: *M* = 9.4 Grade 1: (6.6-8.7 years) *M* = 7.9 Grade 2: 1 year later	Poland	AMAS: .75 LMA: .59 MEA: .71	MLR	Two-factor	.50	
Milovanović & Branovački, 2020[a]	301 primary school children (51.8% males)	Grades 2–4 *M* = 8.79 (.98)	Serbia	LMA: .83 MEA: .80	WLSMV	Two-factor	.62	
Primi et al., 2020[a]	• 150 primary school children (43% males)	Grades 1-2 *M* = 7.1 (.57)	Italy	AMAS: .76 (.78) LMA: .71 (.78) MEA: .66 (.71)	MLM	Two-factor	.67	Strict invariance across gender[b]
	• 223 elementary school students (53% males)	Italian sample: *M* = 6.41 (0.49)	Italy	AMAS: .79 (.83) LMA: .70 (.79) MEA: .71 (.76)	MLM	Two-factor	.77	Strict invariance across languages/ educational
	46%, Italy and 54 % UK	UK sample: *M* = 7.11 (0.66)	UK	AMAS: .74 (.78) LMA: .60 (.74) MEA: .63 (.66)			.75	contexts[b]
Present study	• 450 primary school children (46.3% males)	Grades 3-6 (7-12 years); *M* = 11.19 (1.16)	Spain	AMAS: .77 (.82) LMA: .58 (.72) MEA: .76 (.80)	WLSMV	M2: Two-factor M3: Two-factor with Item 5 in common)	.77 .73	Strong invariance across gender
	• 1054 secondary school children (46% males)	Grades 7-12 (11-19 years); *M* = 15.05 (1.68)		AMAS: .85 (.88) LMA: .75 (.80) MEA: .81 (.84)				Partial strong invariance across educational levels

Note. AMAS: Abbreviated Math Anxiety Scale; LMA: math learning anxiety; MEA: math evaluation anxiety (Testing); CFA: confirmatory factor analysis; EFA: exploratory factor analysis; MLM: mean-adjusted maximum likelihood; ADF: asymptotically distribution-free; WLSMV: weighted least square mean and variance adjusted; ML: maximum likelihood. Age is in years (standard deviation in parentheses).

[a] Studies that administered a modified version of the AMAS.

[b] Models with equality constraints on loadings and residual variances were reported, but not models with equality constraints on item intercepts.

Table 2

Model fit for simple groups models

Model	χ^2(df)	RMSEA [90% CI]	CFI	SRMR	TLI
M1	593.11 (27) ***	.118 [.110, .126]	.938	.067	.917
M2	245.59 (26) ***	.075 [.067, .084]	.976	.043	.967
M3	175.49 (25) ***	.063 [.055, .072]	.983	.037	.976

Note. χ^2(df) = scaled chi-square test (degrees of freedom); RMSEA = scaled root mean square error of approximation; CFI = scaled comparative fit index SRMR = standardized root mean square residual; TLI = scaled Tucker-Lewis Index.

*** $p < .001$

Table 3

CFA standardized factor loadings for the different models

Item	M1	M2		M3	
		LMA	MEA	LMA	MEA
AMAS1	0.31	0.*33*	-	0.33	-
AMAS2	0.76	-	0.79	-	0.81
AMAS3	0.68	0.73	-	0.73	-
AMAS4	0.70	-	0.73	-	0.75
AMAS5	0.61	-	0.64	0.34	0.32
AMAS6	0.74	0.80	-	0.80	-
AMAS7	0.68	0.72	-	0.72	-
AMAS8	0.78	-	0.81	-	0.83
AMAS9	0.65	0.70	-	0.70	-

Table 4

Model fit statistics and indices for evaluating measurement invariance across gender and educational level for model M2

	Freed parameters	χ^2(df)	RMSEA [90% CI]	CFI	SRMR	TLI	$\Delta\chi^2$(df)	ΔRMSEA	ΔCFI
Gender									
Configural		276.01 (52) ***	.076 [.067, .085]	.974	.048	.964			
Thresholds		316.55 (70) ***	.068 [.061, .076]	.971	.048	.971	18.80 (18)	.008	.003
Loadings		292.75 (77) ***	.061 [.054, .069]	.975	.049	.977	5.97 (7)	.007	-.004
Educational level									
Configural		234.26 (52) ***	.068 [.060, .077]	.982	.043	.975			
Thresholds		289.07 (70) ***	.065 [.057, .072]	.978	.043	.977	36.93 (18)*	.003	.004
Thresholds 2	$\tau_{2\text{-}1}$	273.85 (69) ***	.063 [.055, .071]	.979	.043	.979	23.27 (17)	.005	.003
Loadings	$\tau_{2\text{-}1}$	307.70 (76) ***	.064 [.056, .071]	.977	.048	.978	25.21 (7)***	.001	.002
Loadings 2	$\tau_{2\text{-}1}$, λ_1	263.68 (75) ***	.058 [.050, .066]	.981	.044	.982	8.09 (6)	.005	-.002

χ^2(df) = Satorra–Bentler scaled chi-square test (degrees of freedom); RMSEA = scaled root mean square error of approximation; CFI = scaled comparative fit index SRMR = standardized root mean square residual; TLI = scaled Tucker-Lewis index; $\Delta\chi^2$(df) = change in χ^2. ΔCFI = change in CFI; ΔRMSEA = change in RMSEA. *** $p < .001$, * $p < .05$

Table 5. Pearson correlations and HTMT ratio of correlations between AMAS subscales and measures of validity

	Correlation		HTMT ratio	
	MEA	LMA	MEA	LMA
MEA	1		1	
LMA	.58	1	.81	1
Worry (MASYC)	.59	.54	.81	.83
Negative Reactions (MASYC)	.46	.49	.74	.78
General Anxiety (STAIC-T)	.49	.44	.61	.58
Numerical Confidence[a] (MASYC)	.27	.20	.37	.33

Note. [a] Items are reversed in scoring

Table 6

Descriptive statistics of the AMAS items for the total, primary, and secondary school samples

	Total					Primary					Secondary				
Item	*n*	*M*	*SD*	Skewness	Kurtosis	*n*	*M*	*SD*	Skewness	Kurtosis	*n*	*M*	*SD*	Skewness	Kurtosis
AMAS1	1504	1.51	0.87	1.86	3.17	450	1.70	1.05	1.53	1.67	1054	1.42	0.76	1.90	3.30
AMAS2	1504	3.04	1.21	-0.01	-0.91	450	2.83	1.28	0.30	-0.96	1054	3.13	1.17	-0.13	-0.79
AMAS3	1504	2.44	1.20	0.43	-0.77	450	2.41	1.29	0.57	-0.78	1054	2.46	1.16	0.36	-0.79
AMAS4	1504	2.91	1.30	0.05	-1.07	450	2.47	1.31	0.48	-0.91	1054	3.10	1.25	-0.09	-0.95
AMAS5	1504	2.74	1.28	0.21	-1.04	450	2.87	1.43	0.12	-1.33	1054	2.69	1.21	0.21	-0.92
AMAS6	1504	1.57	0.84	1.47	1.70	450	1.34	0.71	2.32	5.73	1054	1.67	0.87	1.23	0.95
AMAS7	1504	1.57	0.90	1.74	2.68	450	1.37	0.77	2.44	6.15	1054	1.65	0.94	1.53	1.92
AMAS8	1504	3.79	1.24	-0.69	-0.65	450	3.76	1.34	-0.67	-0.88	1054	3.80	1.19	-0.69	-0.56
AMAS9	1504	1.69	0.94	1.32	1.11	450	1.50	0.82	1.66	2.20	1054	1.76	0.97	1.19	0.76

Note. LMA subscale items: 1, 3, 6, 7, 9; MEA subscale items: 2, 4, 5, 8

Table 7

Summary of regression analysis predicting LMA and MEA scores from grade, gender and educational level

	Variable	*b*	*SE*	95% CI	β		R^2
LMA							.06
	Educational level	-3.70	-0.52	[-5.39, -2.01]	-0.52	***	
	Grade	-0.24	-0.20	[-0.50, 0.02]	-0.20		
	Gender	-0.86	-0.13	[-1.44, -0.28]	-0.13	**	
	Educational level × Grade	0.61	0.85	[0.32, 0.89]	0.85	***	
	Educational level × Gender	-0.43	-0.06	[-1.13, 0.27]	-0.06		
MEA							.08
	Educational level	-0.50	-0.06	[-2.53, 1.53]	-0.06		
	Grade	0.45	0.31	[0.14, 0.77]	0.31	**	
	Gender	-1.26	-0.16	[-1.96, -0.56]	-0.16	***	
	Educational level × Grade	-0.06	-0.07	[-0.41, 0.28]	-0.07		
	Educational level × Gender	-0.81	-0.10	[-1.64, 0.02]	-0.10		

Note. *** $p < .001$, ** $p < .01$

Appendix A

Items in the original and adapted versions of the AMAS

Item	Original	Spanish
1	Having to use the tables in the back of a math book	Teniendo que usar las tablas del final del libro de matemáticas
2	Thinking about an upcoming math test 1 day before	Pensando en el próximo examen de matemáticas el día antes de hacerlo
3	Watching a teacher work an algebraic equation on the blackboard	Viendo a un profesor/a haciendo una operación matemática complicada en la pizarra
4	Taking an examination in a math course	Haciendo un examen en una asignatura de matemáticas
5	Being given a homework assignment of many difficult problems that is due the next class meeting	Si te mandan deberes con muchos problemas difíciles para entregar en la próxima clase
6	Listening to a lecture in math class	Escuchando una lección en la clase de matemáticas
7	Listening to another student explain a math formula	Escuchando a otro alumno/a explicar un problema de matemáticas
8	Being given a “pop” quiz in math class	Si te hacen un examen sorpresa en la clase de matemáticas
9	Starting a new chapter in a math book	Empezando una lección nueva del libro de matemáticas

Appendix B

Table A

Means, standard deviations and percentiles for the LMA and MEA scores by grade

		Males										Females									
					Percentiles										Percentiles						
	Grade	*n*	*M*	*SD*	5	10	25	50	75	90	95	*n*	*M*	*SD*	5	10	25	50	75	90	95
LMA																					
	3	61	7.8	2.6	5.0	5.0	6.0	7.0	10.0	12	13.0	54	9.3	3.2	5.0	6.0	6.8	9.0	11.0	14.0	16.3
	4	47	8.4	3.1	5.0	5.0	6.0	8.0	10.0	13.0	14.2	59	8.2	2.7	5.0	5.0	6.0	8.0	10.0	12.0	13.0
	5	52	8.3	2.7	5.0	5.0	6.0	8.0	10.0	12.7	14.4	73	9.3	3.6	5.0	5.0	6.0	9.0	12.0	14.0	17.0
	6	48	7.1	1.9	5.0	5.0	5.3	7.0	8.8	9.0	11.1	56	7.9	2.4	5.0	5.0	6.0	7.0	9.8	11.3	13.0
	7	65	6.7	2.1	5.0	5.0	5.0	6.0	8.0	10.0	11.7	93	8.8	3.0	5.0	5.4	6.0	8.0	10.5	13.0	15.0
	8	88	7.8	2.5	5.0	5.0	6.0	7.5	9.0	11.0	14.1	112	9.3	3.3	5.0	5.0	6.0	9.0	12.0	14.7	16.0
	9	85	8.3	3.1	5.0	5.0	6.0	7.0	11.0	13.0	14.7	98	9.7	3.8	5.0	5.0	7.0	10.0	12.0	15.0	18.0
	10	95	8.1	2.9	5.0	5.0	6.0	7.0	10.0	12.0	13.0	104	8.9	3.7	5.0	5.0	6.0	8.0	11.8	15.0	16.0
	11	89	9.7	3.6	5.0	5.0	7.0	9.0	12.5	15.0	15.5	101	10.5	3.2	5.1	6.2	8.0	10.0	13.0	15.0	16.0
	12	63	8.9	3.2	5.0	5.0	6.0	9.0	11.0	13	15.8	61	10.1	3.7	5.0	5.0	7.0	10.0	13.0	15.8	17.0
MEA	3	61	10.4	4.0	4.0	5.2	7.0	10.0	14.0	16.0	17.0	54	11.5	4.2	4.8	6.0	8.0	11.5	15.0	18.0	18.3
	4	47	11.3	4.4	5.0	5.8	8.0	10.0	15.0	17.2	19.2	59	12.1	4.1	5.0	7.0	9.0	12.0	15.0	18.0	20.0
	5	52	11.8	3.7	5.0	6.3	9.3	12.0	14.0	16.7	18.0	73	13.7	4.3	6.7	7.0	11.0	14.0	17.0	19.0	20.0

6	48	11.6	3.7	5.5	6.0	8.0	12.0	14.8	17.0	17.0	56	12.5	3.5	7.0	7.0	10.0	12.0	15.0	18.0	19.0
7	65	10.6	4.0	5.0	5.6	7.0	10.0	13.0	16.4	17.7	93	13.0	3.6	6.0	8.0	10.0	14.0	16.0	17.6	19.0
8	88	11.3	4.0	4.5	6.0	8.3	11.0	15.0	17.0	18.0	112	13.4	3.7	6.0	7.3	11.0	14.0	16.0	18.0	19.0
9	85	10.6	4.3	4.0	5.0	7.0	10.0	14.0	16.4	18.0	98	13.4	3.3	7.0	8.9	11.8	14.0	15.3	17.1	19.0
10	95	11.6	3.3	6.0	7.0	9.0	12.0	14.0	16.0	17.0	104	13.0	3.8	5.3	7.5	11.0	13.0	16.0	18.0	19.0
11	89	13.2	3.5	6.5	8.0	11.0	14.0	15.5	18.0	18.5	101	15.3	3.3	9.0	10.0	14.0	16.0	18.0	19.0	19.0
12	63	12.4	3.3	7.0	7.4	11.0	12.0	15.0	17.0	17.8	61	13.8	2.8	9.0	10.0	12.0	14.0	16.0	17.0	18.9

Note. LMA = math learning anxiety subscale; MEA= math evaluation anxiety subscale