Research Reports

Norms and Validation of the Online and Paper-and-Pencil Versions of the Abbreviated Math Anxiety Scale (AMAS) For Polish Adolescents and Adults

Krzysztof Cipora*[ab], Klaus Willmes[c], Adrianna Szwarc[d], Hans-Christoph Nuerk[aef]

[a] Department of Psychology, University of Tuebingen, Tuebingen, Germany. **[b]** Copernicus Center for Interdisciplinary Studies, Cracow, Poland. **[c]** Department of Neurology, University Clinic RWTH Aachen University, Aachen, Germany. **[d]** Institute of Psychology, Jagiellonian University, Cracow, Poland. **[e]** Leibniz-Institut für Wissensmedien, Tuebingen, Germany. **[f]** LEAD Graduate School & Research Network, University of Tuebingen, Tuebingen, Germany.

Abstract

The Abbreviated Math Anxiety Scale (AMAS) is one of the most popular instruments measuring math anxiety (MA). It has been validated across several linguistic and cultural contexts. In this study, we investigated the extent of administration method invariance of the AMAS by comparing results (average scores, reliabilities, factorial structure) obtained online with those from paper-and-pencil. We administered the online version of the AMAS to Polish students. Results indicate that psychometric properties of the AMAS do not differ between online and paper-and-pencil administration. Additionally, average scores of the AMAS did not differ considerably between administration forms, contrary to previous results showing that computerized measurement of MA leads to higher scores. Therefore, our results provide evidence for the usefulness of the AMAS as a reliable and valid MA measurement tool for online research and online screening purposes across cultures and also large similarity between administration forms outside an American-English linguistic and cultural context. Finally, we provide percentile and standard norms for the AMAS for adolescents and adults (in the latter case for both online and paper-and-pencil administration) as well as critical differences for the comparison of both subscales in an individual participant for practical diagnostic purposes.

Keywords: math anxiety, Abbreviated Math Anxiety Scale, online survey, validation, norms

Math Underachievement – Its Causes and Consequences

Poor numeracy may even be more harmful for an individual than poor literacy (Butterworth, Varma, & Laurillard, 2011). Considerable large-scale economic consequences of math underachievement have also been reported (OECD, 2010). Therefore, the cognitive and emotional causes of impaired numeracy are of major interest both theoretically and practically. Regarding emotional aspects, math anxiety (MA) is probably the most prominent and important variable that negatively correlates with math achievement. It has been extensively investigated since the 1960's (see Dowker, Sarkar, & Looi, 2016 for review). A sizeable negative relationship between MA and math performance is evident in previous PISA (*Programme for International Student Assessment*) studies

across virtually all countries involved in the program (see Lee, 2009 for PISA 2003). PISA 2012 (OECD, 2013) showed that 14% of variance in math performance was accounted for by students' MA levels. In Poland, which is of particular importance for this study, it was 22% and this score was one of the highest among countries involved in the PISA programme. Nevertheless, MA levels seem to differ considerably between countries. Importantly, the same study revealed an increase in average MA levels in several countries compared to PISA 2003, although its relationship with math achievement did not change considerably. Thus, the problems for societies associated with MA seem to be growing rather than diminishing on a global level. Therefore, current research should address the assessment of MA in order to measure it online and on paper-and-pencil versions as reliably and validly as possible. The goal of this study is to evaluate the validity and reliability of an online version of the Abbreviated Math Anxiety Scale (AMAS; Hopko, Mahadevan, Bare, & Hunt, 2003) outside the English-American Culture (in a Slavic culture) and to provide norms and critical differences for practical diagnostic assessment.

Math Anxiety – A Specific Instance of Anxiety

Math anxiety refers to a wide range of negative emotional responses related to math and situations involving math (Ashcraft & Ridley, 2005). Negative emotional responses vary from apprehension to dread, and may appear in both academic situations involving math as well as in everyday life situations that require using math (e.g. financial transactions).

Importantly, MA cannot be subsumed under general anxiety or test anxiety. The evidence for its specificity comes from psychometric data both in adults (e.g., Ashcraft & Ridley, 2005), and in children (Carey, Hill, Devine, & Szűcs, 2017). Converging evidence comes also from physiological (e.g. Ashcraft, 2002) and neuroscientific studies (Artemenko, Daroczy, & Nuerk, 2015 for discussion and some controversies), which can typically dissociate math and general anxiety.

However, MA cannot be subsumed under poor math skills either. It negatively correlates with math achievement, although only moderately (averaged correlation - .27; Ma, 1999 for meta-analysis). In line with this moderate relationship between MA and math ability, neurocognitive activation differences between high and low MA participants can be found even when math ability is matched (Pletzer, Kronbichler, Nuerk, & Kerschbaum, 2015). This confirms that MA and math skills need to be considered as independent concepts and not just as cognitive and affective sides of the same coin. The relationship between MA and math achievement seems to be reciprocal; poor math skills increase MA, and MA mitigates math performance (Carey, Hill, Devine, & Szűcs, 2016). Therefore, it is not surprising that interventions aimed at reducing MA lead to improvement in math achievement (Hembree, 1990 for meta-analysis). Importantly, the relation between MA and math skills is not unitary, but may be additionally qualified by at least three factors.

i. Very recently, Carey, Devine, Hill, and Szűcs (2017b) proposed that this relationship may be influenced by the individual's anxiety profile (i.e., the relationship between math anxiety and other types of anxiety) rather than MA levels only.

ii. Observed performance differences between high- and low-anxiety individuals largely depend on testing conditions; the relationship is stronger in formal testing, especially when individuals work under time pressure (Ashcraft & Faust, 1994).

iii. There is evidence for moderate genetic and non-shared environmental influences on MA (Wang et al., 2014), which might qualify the relation to math skills.

In sum, MA is a valid construct and cannot be reduced to anxiety in general or solely attributed to poor math skills. Studying this phenomenon is instructive for educational psychology and may contribute to our understanding of how anxiety affects (numerical) cognition in general (see e.g., Suárez-Pellicioni, Núñez-Peña, & Colomé, 2016 for a review).

Gender Differences in Math Anxiety

Data on gender differences in MA are inconsistent across cultures, age groups and studies. The majority of studies suggests that females (both adults and children) tend to have higher MA than males. Such differences (i.e., women revealing higher levels than men) have been reported in the U.S. (see e.g., Hembree, 1990), the U.K. (Hunt, Clark-Carter, & Sheffield, 2011), Australia (Gyuris & Everingham, 2011), and outside Anglo-Saxon countries in Italy (Primi, Busdraghi, Tomasetto, Morsanyi, & Chiesi, 2014), Spain (Núñez-Peña, Suárez-Pellicioni, Guilera, & Mercadé-Carranza, 2013), and Poland (Cipora, Szczygieł, Willmes, & Nuerk, 2015a), for instance. In line with these findings, females sometimes performed worse on math tests than cognitive capacity-matched males, and this poorer performance was attributed to lower MA (Devine, Fawcett, Szűcs, & Dowker, 2012). However, some studies report no gender differences (see e.g., Baloğlu, 2003; see also Gyuris & Everingham, 2011) or even higher MA in males (Devine et al., 2012 for a review). For instance, males scored higher in math anxiety in an Iranian sample (Vahedi & Farrokhi, 2011). These data seem to suggest that gender differences observed in Western cultures may not be universal.

Similarly, cross-cultural variability in gender differences in MA was present in PISA 2012 (OECD, 2013). The effect size of this difference varied from $d = 0.51$ in favor of boys in Denmark and Switzerland (i.e., boys being characterized by lower MA than girls) to $d = 0.19$ in favor of girls in Qatar and Jordan (i.e., girls being characterized by lower MA). In Poland, the significant gender gap was small in size ($d = 0.11$ in favor of boys). Note, however, that PISA studies adolescents (15 year olds), thus the results may not be representative for adults. Indeed, a study on Polish adults (Cipora et al., 2015a) showed a much larger gender gap in MA ($d = 0.61$ in favor of males) than the PISA data of the same country (for possible reasons, see below).

In sum, because of the inconsistent findings across cultures, age groups and studies, we need reliable data from different cultures to better characterize cultural variation of MA and its gender differences. These data should come from cultures other than large Western cultures, in which the vast majority of studies are done, because of the strong cross-cultural differences reviewed above.

Structure of Math Anxiety

The structure of MA is a matter of scientific debate: 2-factor, 3-factor, and 6-factor models have been proposed and will be shortly outlined here:

2-factor models: Usually MA is believed to comprise two factors: anxiety related to using math in everyday situations, and anxiety related to being evaluated in math (Suinn & Edwards, 1982). Sometimes, the first factor is related to anxiety in the context of learning math while the other is more related to test anxiety. The two MA factors are highly correlated (Hopko, 2003). The two-factor model is the most popular and common model of MA.

3-factor models: A three-component model was proposed by Alexander and Martray (1989) with the factors (1) Math test anxiety, (2) Numerical Task Anxiety, and (3) Math Course Anxiety. Three factors also account best for the pattern of scores in the Mathematics Anxiety Scale – UK (MAS-UK) that was developed for the British population (Hunt et al., 2011). These factors were termed (1) Maths Evaluation Anxiety, (2) Everyday / Social Maths Anxiety, (3) Maths Observation Anxiety. Sometimes abstraction anxiety (anxiety related to abstract mathematical content) is also considered to be another component of MA (Ma & Xu, 2004).

6-factor models: Analyzing several tools measuring MA, Kazelskis (1998) described 6 oblique factors of math anxiety involving: (1) Mathematics Test Anxiety, (2) Numerical Anxiety, (3) Negative Affect Toward Mathematics, (4) Worry, (5) Positive Affect Toward Mathematics, and (6) Mathematics Course Anxiety. This factor structure did not hold in the light of subsequent research. However, positive affect towards mathematics is sometimes considered as a component of math anxiety (Bai, 2011). The questionnaire MAS-R developed by this latter author was shown to reliably measure two independent aspects of math anxiety – positive and negative. Based on a confirmatory factor analysis of the MARS30-Brief scale, Pletzer et al. (2016) conclude that a simple distinction between numerical anxiety and testing anxiety does not satisfactorily fit the data. Best fit was obtained by a model consisting of six factors (1) Evaluation Anxiety 1 – proper - Taking Math Exam, (2) Evaluation Anxiety 2 – Thinking of Upcoming Exam, (3) Learning Math Anxiety, (4) Everyday Numerical Anxiety, (5) Performance Anxiety, (6) Social Responsibility Anxiety.

In sum, the factorial structure of math anxiety is relatively variable and different structural proposals currently exist in the literature, however, most researchers agree that it is not a unidimensional construct.

AMAS – Abbreviated Math Anxiety Scale

One of the more recent instruments for measuring MA is the Abbreviated Math Anxiety Scale (AMAS) developed by Hopko et al. (2003). Since we are using the AMAS in this study, we outline its diagnostic properties in more detail.

Overall Characteristics

AMAS is a 9-item questionnaire, characterized by very good psychometric properties. It takes less than 5 minutes to administer, making it very convenient as a screening tool and for research purposes. Apart from the total score, AMAS comprises two scales: (1) anxiety related to learning math (*Learning*), and (2) anxiety related to being tested in math (*Testing*). This structure was confirmed in several samples using both exploratory and confirmatory factor analyses (e.g., Hopko et al., 2003; Cipora et al., 2015a, 2015b).

Reliability

AMAS reliability is mostly around or above .80 and was measured by means of internal consistency (Cronbach's alpha and ordinal alpha) and test-retest reliabilities. In the first study, Hopko et al. (2003) reported its good internal consistencies (≥ .85) and test-retest reliability (≥ .78). These estimates were similar in Iranian (Cronbach's alpha ≥ .75; Vahedi & Farrokhi, 2011), Italian (Cronbach's alpha ≥ .80; Primi et al., 2014), Polish (Cronbach's alpha ≥ .78; test-retest reliability ≥ .59; ordinal alpha ≥ .84; Cipora et al., 2015a, 2015b), German (Cronbach's alpha for total score = .92; Dietrich, Huber, Moeller, & Klein, 2015), and Spanish (Cronbach's alpha ≥ .87; Brown & Sifuentes, 2016) language versions. Reliability holds both for adults, high-schoolers (Cronbach's alpha for Italian high-schoolers ≥ .81; Primi et al., 2014 and ≥ .76 for Polish secondary- and high

schoolers; Cipora et al., 2015b) and primary schoolers 8-11-year-olds (Cronbach's alpha in Italian sample ≥ .64; Caviola, Primi, Chiesi, & Mammarella, 2017). In a modified form AMAS was shown to be reliable for measuring math anxiety in British 8-13-year-olds (Cronbach's alpha > .74; ordinal alpha > .83; Carey, Hill, et al., 2017). It was also used in studies on English speaking younger adolescents (11 years and older, however reliability estimates for this age group were not provided; Devine et al., 2012). Adequate reliability estimates were also reported for the modified AMAS administered to Australian students (Cronbach's alpha ≥ .83; Gyuris & Everingham, 2011; Gyuris, Everingham, & Sexton, 2012).

In sum, while AMAS reliability can be considered satisfactorily for a personality questionnaire in general, there are some cultural variations and especially some lower reliabilities for younger children.

Validity

In general, AMAS has been shown to possess convergent and discriminant validity: It correlates positively with other MA measurement scales (MARS-R; .85), but only moderately with computer anxiety (.32), test anxiety (.52) and trait anxiety (.28) (Hopko et al., 2003). Factor analyses indicate sufficient construct validity of the two-factor structure both in adults (e.g., Hopko et al., 2003; Cipora et al., 2015a) and adolescents (Carey, Hill, et al., 2017; Caviola et al., 2017; Cipora et al., 2015b). Similar indices of construct, convergent and discriminant validities were reported in other language versions, including Farsi (e.g., .61 correlation with statistics anxiety measure; Vahedi & Farrokhi, 2011), Italian (e.g., .57 correlation with test anxiety; Primi et al., 2014; .32 correlation with the physiological anxiety subscale; Caviola et al., 2017) and Polish (e.g., .33 correlation with trait anxiety; Cipora et al., 2015a, 2015b). Carey, Hill, et al. (2017) in a large scale study by means of item-level factor analyses showed that (slightly modified) AMAS items loaded on a separate factor among other factors representing other types of anxiety (varied aspects of general anxiety and test anxiety).

AMAS was also shown to have expected patterns of correlations with more general differential psychology constructs, namely temperamental traits (e.g., .48 correlation with Emotional Reactivity; no correlation with Sensory Sensitivity nor Activity), as well as attitudes towards math and other school subjects (e.g., -.50 correlation with liking math and .12 correlation with liking humanities; Cipora et al., 2015a, 2015b).

Importantly, the AMAS structure was shown to be gender-invariant, which was found both for the original English version (Hopko et al., 2003) as well as for other language versions including Farsi (Vahedi & Farrokhi, 2011), Italian (Caviola et al., 2017; Primi et al., 2014) and Polish (Cipora et al., 2015a).

The above results suggest that the AMAS can validly measure MA in various linguistic and cultural contexts. Therefore, the AMAS is also potentially suitable for (cross-cultural) online studies. However, with the rise of online assessment, validity has become a crucial question. In the next section, we will discuss commonalities and specificities of online questionnaire measurement employed in this study compared to conventional assessment.

Online Questionnaire Measurement – Commonalities and Specificities Compared to Paper-and-Pencil Administration

Online psychological measurement is becoming more and more popular. Advantages and disadvantages of online studies are extensively discussed by Reips (2002; see in particular their Table 1, p. 245; see also Reips, Buchanan, Krantz, & McGraw, 2015 for guidelines on online data interpretation). Online questionnaires are

especially valuable for short surveys that take no more than 10-15 minutes, so that it is easy to recruit participants and the dropout rate is low (Reips, 2002).

It seems that the mode of administration does not necessarily influence psychometric properties of measurement instruments. It also seems that differences between administration forms, observed when computerized measurement was first introduced, disappear over time and with familiarity with computers and the Internet. Close correspondence was demonstrated between paper-and-pencil versus computerized tests, but both tests were administered in the lab (see: Gwaltney, Shields, & Shiffman, 2008, for a meta-analysis of therapy outcome measurement; Booth-Kewley, Edwards, & Rosenfeld, 1992 for computer anxiety; King & Miles, 1995 for desirable social responding, sensitivity of equity and self-esteem). Similarly, strong correspondence was observed between results of online surveys and questionnaires distributed via traditional mail, e.g., for personality measurement outcomes (Pettit, 2002) or for several aspects of physical health and well-being (Ritter, Lorig, Laurent, & Matthews, 2004). Nevertheless, these authors explicitly recommend collecting normative data for each administration form.

It seems that under some provisions like careful control for inclusion criteria, large samples and post-hoc data filtering to exclude randomly responding participants, online testing can be used in psychometric studies. It also seems that in many, but possibly not all instances, results from online and paper-and-pencil administration can provide equivalent results.

Applications of Computerized / Online Surveys for Math Anxiety Measurement

Ashcraft and Faust (1994) administered a paper-and-pencil version of MARS to half of their participants, and the other half received a computerized form of the same test, with the only difference between the two versions being that in the computerized form, items were presented one by one. Unexpectedly, participants who received the computerized form scored significantly higher (d = 0.62) than those who received the paper-and-pencil form. The authors ascribe this difference to the fact that using a computer evoked even more anxiety in individuals who normally are math anxious. Nevertheless, despite differences in means, observed effects of gender differences and correlation with math problem solving efficiency were similar for both administration modes. However, this study was conducted in the early 1990's, when using computers was much less habitual than it is today – participants may have been more excited to be assessed via computer questionnaires than today. Therefore, we believe that it is important to reassess the limitations of online assessment validity, i.e. claim that MA scores obtained in computerized procedures will differ from paper-and-pencil administration.

It is important to note that some studies have administered the AMAS in an online version. Jones, Childers, and Jiang (2012) administered the AMAS twice; the first time as a paper-and-pencil and the second time as an online survey. The reported test-retest reliability was very high (r = .91), providing initial evidence both for test-retest reliability and AMAS suitability for online measurement. However, validation of online measurement was not the objective of that study, and no other information was provided – thus, we do not know, whether the higher values of Ashcraft do still hold, because high correlations are also possible, when all participants score a little bit higher in the online version. Ferguson, Maloney, Fugelsang, and Risko (2015) also administered the AMAS in an online format as part of a larger online battery aimed at investigating relationships between MA and spatial abilities. The study also revealed an expected pattern of correlations, however the validity of the AMAS online as compared to its paper-and-pencil version was not evaluated.

Objectives of the Study

The AMAS has been previously used in online administration mostly in English-speaking samples. Here, we evaluated reliability and validity of the AMAS online version in a different linguistic and cultural background, because of cross-cultural differences reported for offline versions of math anxiety assessment. However, equivalence with paper-and-pencil administration cannot be taken for granted as the above review showed – so the construct validity and the norms of the online version will be compared to a paper-and-pencil version. Moreover, the comparison of online assessment with paper-and-pencil assessment allows us to test whether early claims of differences between MA scores between computer and paper-and-pencil administration still hold, when online assessment and computer use has become much more common (Ashcraft & Faust, 1994).

Finally, for investigating any individual characteristics, both researchers and practitioners should be equipped with adequate normative data. Thus, following up on the initiative by Caviola et al. (2017), who published AMAS percentile norms for Italian primary school children, we prepared norms for the Polish AMAS based on data we collected up to date (*n* = 2057) and we hope this initiative will be taken up by other researchers.

Method

Participants

A total of 615 participants (418 female and 197 male), mean age 22.0 years (*SD* = 3.9; range 17-50) participated in the study.

Participants were recruited via email circulated through university mail at several universities in Poland in various cities (Kraków, Lublin, Warszawa, Toruń, Wrocław) as well as by sharing a link via social media. In the announcement, we invited individuals to participate in a study, and if possible, to share the invitation with other potentially interested individuals. In the invitation it was stated that the whole procedure will take about 3 minutes and that the scale comprises only nine items. It also stated that the aim of the study was to prepare the validation of a questionnaire aimed at measuring a specific kind of anxiety in a Polish sample. All receivers were free to ignore the email and did not receive any compensation for their participation. All participants received the same link, therefore ensuring data anonymization.

Most participants were university students; however, some participants recruited via social media were also employees. They represented a varied range of fields of studies / occupations: Psychology (*n* = 380), STEM (science, technology, engineering, and mathematics; *n* = 141; comprising: engineering - varied specializations, computer science, architecture etc.), neurobiology / neuropsychology / cognitive science (*n* = 54), humanities and pedagogy (*n* = 27; comprising: pedagogy, philosophy, literature etc.), other (*n* = 11) and two individuals did not answer this question. (In the shared dataset one can access the study sample composition by gender. This can be done by running the specified parts of the shared R code, see Supplementary Material.)

Materials and Procedure

The online version of the AMAS questionnaire was identical to the paper-and-pencil version in terms of content and scale formation (see Cipora et al., 2015a). The only difference was that a nickname, age, gender, and field of study sections were presented at the bottom of the website.

The form started with an instruction stating that the participant will see some statements related to learning math. She or he was asked to mark the level of anxiety it evokes/would evoke in her/him. Responses were given on a 5-point Likert scale. Scale extremes were labelled with *mild anxiety* and *strong anxiety*.

The Google Documents form was used in the default graphic. All items were marked as obligatory. The entire questionnaire fit on one page. After clicking the link, participants were redirected to the Google Documents form. The whole procedure lasted approximately 3 minutes.

The theoretical range of sum scores across all nine items (AMAS total score) was 9-45, whereas for the *Learning* and *Testing* scales it was 5-25 and 4-20 respectively.

Data Analysis

In the present work, we examined the reliability and construct validity of the AMAS questionnaire. We compared the AMAS results obtained from this online study with the previous findings for paper-and-pencil administration to evaluate commonalities and differences between online and paper-and-pencil administration.

To estimate AMAS reliability, we measured Cronbach's alpha, which is one of the most popular reliability measures. Nevertheless, feasibility of this reliability measure has been strongly criticized recently. Critics point out that it is computed using Pearson correlation coefficients among items, which may be biased when the raw data is not continuous and the number of different item-scale values is small, which is the case for Likert-type responses. Therefore, Cronbach's alpha may underestimate reliability, and this problem can be even more severe for scales comprising only a small number of items like the AMAS (Yang & Green, 2011). Another source of problems may originate from non-normal distributions of both true scores and error scores (Sheng & Sheng, 2012). Ordinal alpha, computed using the polychoric correlation coefficient, can be an alternative that addresses these issues. Polychoric correlations take into account the fact that the observed data (i.e., Likert-type data) are discrete manifestations of a continuous latent construct (Zumbo, Gadermann, & Zeisser, 2007).

We investigated the AMAS factor structure by means of confirmatory factor analysis. Namely, we examined whether the factor structure of AMAS observed in the previous study can be replicated in an independent dataset collected with a different administration form. Additionally, we conducted an exploratory factor analysis (Appendix A) with oblique rotation, because the theoretical expectation is that different aspects of MA are inherently correlated.

Lastly, we compared the AMAS results obtained via online administration with the data obtained in a previous study (Cipora et al., 2015a) based on the paper-and-pencil setup of the AMAS. In the case of reliability estimates and correlations, we also calculated 95% confidence intervals to compare reliabilities obtained via online and paper-and-pencil administrations.

Analyses were conducted in R (R Core Team, 2017) and SPSS 24. The confirmatory factor analysis was conducted in AMOS 24 software (Arbuckle, 2016).

Data Availability

Raw data along with R script and SPSS syntax we used for analyses can be found in the Supplementary Material.

Results

Descriptive Statistics

The average AMAS total score was 20.96 (*SD* = 6.22), slightly below the theoretic scale midpoint of 27. The distribution deviated significantly from normality as assessed with the Shapiro-Wilk test, *W*(615) = 0.98, *p* < .001. Skewness was 0.43 (*SE* = 0.10) and kurtosis was -0.24 (*SE* = 0.20). Skewness divided by its standard error falls outside the ±2 range, so the distribution may be considered to be skewed. However, with respect to kurtosis, there was no significant departure from normality. The average score for the *Learning* scale was 7.59 (*SD* = 2.80). It was close to the theoretical scale minimum of 5. A subgroup of 171 participants (27.8% of the sample) scored the scale minimum, 331 participants (53.8%) scored 7 or less. The distribution significantly deviated from normality, *W*(615) = 0.86, *p* < .001. Skewness was 1.23 (*SE* = 0.10) and kurtosis was 1.63 (*SE* = 0.20). Both skewness and kurtosis divided by their respective standard error fell outside the ±2 range, indicating strong deviation from normality. In case of the *Testing* scale the average score was 12.97 (*SD* = 3.94), above the theoretical scale midpoint of 12. The distribution also deviated significantly from normality, *W*(615) = 0.97, *p* < .001. Skewness was -0.17 (*SE* = 0.10) and kurtosis was -0.85 (*SE* = 0.20). Skewness divided by its SE falls within ±2 range, so there is no major deviation from normality for this parameter, although this is not the case for the kurtosis estimate. Distributions of results for the scale total as well as for both scales are depicted in Figure 1.
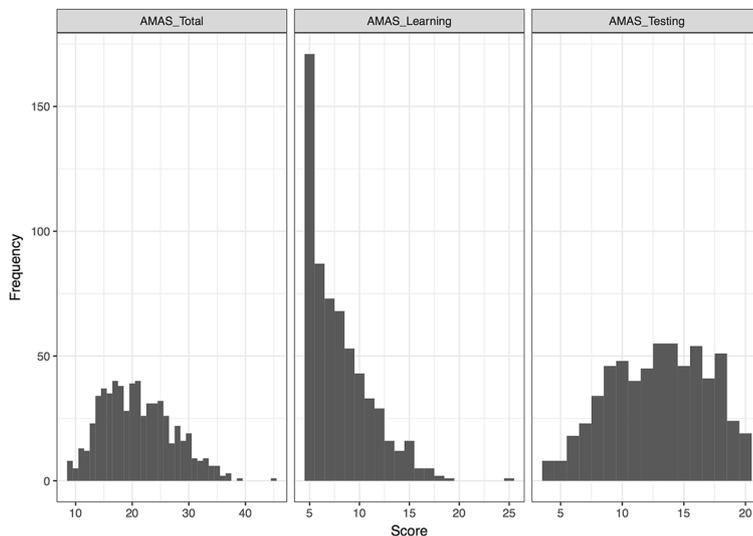
Figure 1. Distributions of AMAS online. Total score (Panel A; theoretical range from 9 to 45), AMAS *Learning* scale (Panel B; theoretical range from 5 to 25) and AMAS *Testing* scale (Panel C; theoretical range from 4 to 20).

Descriptive statistics for all items individually are also reported in Table 1. By conducting reliability analysis with supplementary R code one can access a table containing proportion of each response alternatives to each item. By using another part of the code one may generate a plot presenting frequencies of given response alternatives for each item. The Pearson correlation between both scales was .55. Correlation between the *Learning* scale and the total score was .85, and the correlation between the *Testing* scale and the total score was .91.

For the total score as well as for both sub-scales there were no differences in variance between genders (Levene's tests all ps > .139). Female participants scored higher than males. In females, the average total score was 21.62 (SD = 6.33) and for males it was 19.54 (SD = 5.73), t(613) = -3.91, p < .001; d = 0.34. Noteworthy, when the analysis was conducted separately within each field of study, the differences were present in the STEM (i.e., science, technology, engineering, and mathematics) group only (d = 0.53). By running the supplementary R code one may access more detailed statistics.

In the *Learning* scale, the mean score for females was 8.17 (SD = 3.22), and for males it was 7.59 (SD = 2.80), t(613) = -2.16, p = .031, d = 0.19. In the *Testing* scale, females scored 13.45 (SD = 3.95) whereas males scored 11.95 (SD = 3.73), t(613) = -4.48, p < .001, d = 0.39.

Reliability

Cronbach's Alpha

Cronbach's alpha coefficient reliability estimates were .84, .72 and .85 for the AMAS total, *Learning,* and *Testing* scales, respectively. Coefficients did not increase after any item was excluded. Average inter-item correlation was .37, .34, and .59 for the total score, *Learning*, and *Testing* scales, respectively. Item characteristics are summarized in Table 1.

Table 1

AMAS Item Characteristics and Factor Loadings From Confirmatory Factor Analysis (n = 615).

| Item | Item content | Scale | *M* | *SD* | Corrected[a] item total correlation | | | CFA, squared multiple correlation | |
					total	*Learning*	*Testing*	*Learning*	*Testing*
1	Using tables	*Learning*	1.51	0.88	.38	.36	.32	.19	-
2	Test one day before	*Testing*	3.03	1.23	.71	.51	.75	-	.75
3	Watching teacher's work	*Learning*	1.67	0.96	.57	.57	.45	.52	-
4	Math exam	*Testing*	3.64	1.15	.61	.39	.70	-	.61
5	Homework	*Testing*	2.65	1.18	.64	.52	.59	-	.53[b]
6	Attending lecture	*Learning*	1.64	0.92	.48	.53	.36	.50	-
7	Other student explaining Math	*Learning*	1.60	0.88	.43	.49	.32	.39	-
8	Pop quiz	*Testing*	3.65	1.17	.65	.42	.73	-	.68
9	New chapter	*Learning*	1.57	0.86	.52	.46	.45	.44	-

[a]In the case of correlation with the total score and the scale a given item is assigned to, corrected item-total correlations are provided (i.e., correlations with the scale score excluding the given item); in case of correlations with the other scale, bivariate correlations between the item and the scale score are presented. [b]Allocated to scale with higher loading.

Ordinal Alpha

In order to estimate ordinal alpha, we used the method proposed by Gadermann, Guhn, and Zumbo (2012). Ordinal alpha was .88, .80, and .88 for the total scores, *Learning,* and *Testing* scales, respectively. Reliabilities did not increase after any item was excluded.

Construct Validity: AMAS Factor Structure – Confirmatory Factor Analysis

In the second step of analysis, we performed a confirmatory factor analysis on our data. We examined the fit of the model that attained the best fit for our paper–and-pencil data (see Cipora et al., 2015a, 2015b). The model

reflects the structure of the AMAS postulated by Hopko et al. (2003), with the only exception being item no. 5 (Homework), which was modeled to contribute to both latent variables. Nevertheless, the model comprising single loading for this item also reached acceptable fit to the data.

The multivariate normality assumption was violated (multivariate kurtosis = 26.73; critical ratio = 23.55). Therefore, similar to the paper-and-pencil data, we used an asymptotically distribution-free (ADF) method. This method is suited for modeling Likert-type data and may be considered equivalent to estimation based on polychoric correlations. For the same reason, CMIN/DF measures of model fit are not reported, since they are sensitive to violations of the normality assumption (Bedyńska & Książek, 2012). The model, together with standardized path coefficients, is presented in Figure 2. All parameters differed significantly from 0 which means that all factor loadings were significant. Noteworthy, the same model reached acceptable fit for both female and male participants groups separately.
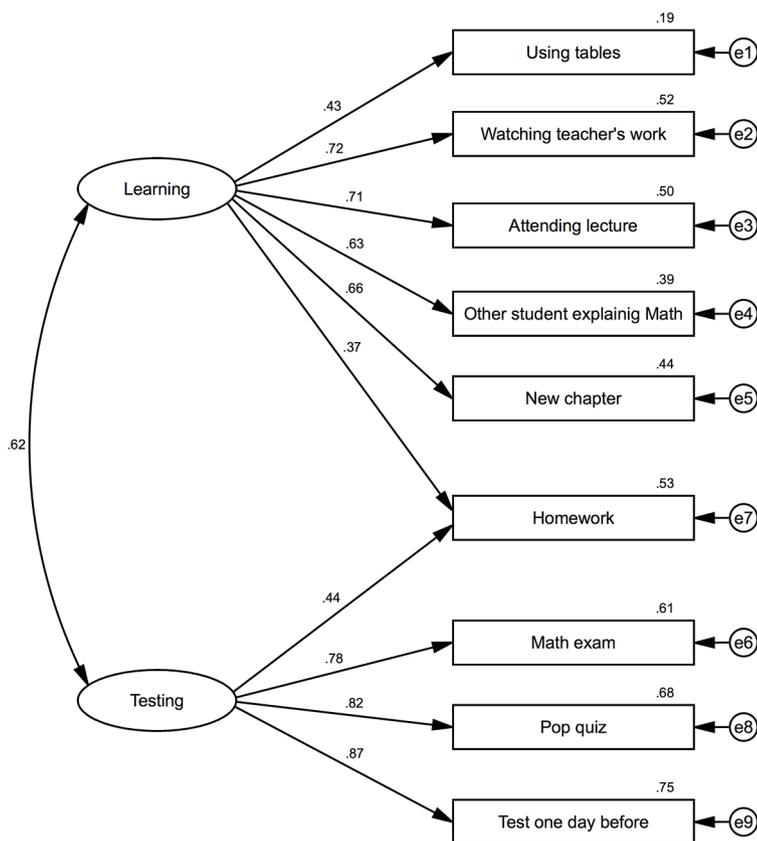
Figure 2. Confirmatory factor analysis on AMAS online data. Indices of the model fit are presented in the main text. Standardized coefficients are presented along with the respective paths. Internal structure of the AMAS online data is highly similar to paper-and-pencil data. Variables labelled with e1, e2, etc. denote respective error terms.

Apart from Item 1 (Using tables) and Item 5 (Homework), all path coefficients were at an acceptable level of > .60. Squared multiple correlations for each item are presented in Table 1. Apart from Item 1, all factor loadings are close to or above .40. The fit of the model was good (RMSEA = .079; 90% confidence intervals .065-.093; AGFI = .895).

In the next step, we investigated invariance of the factor structure across genders. As AMOS 24 does not allow multi-group analysis when the ADF method is used, we conducted another analysis using the maximum likelihood method. We estimated two nested models, unconstrained and the model with the same measurement weights between genders. Formal comparison of χ^2 values revealed that the unconstrained model fit the data better than the constrained one ($p = .001$). On the other hand, using χ^2 statistic to evaluate overall model fit is problematic (i.e., in case of large sample sizes the χ^2 statistic almost always rejects a model; Jöreskog & Sörbom, 1993). Using this method for model comparison raises the same problems (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Overall fit of both the constrained and unconstrained model estimated by means of RMSEA was .056 for the unconstrained and .57 for the constrained model, and confidence intervals were .046 - .067 and .047 - .067 respectively.

Format Commonalities and Differences: Online Versus Paper-and-Pencil Administration

In order to further evaluate the online version of the AMAS questionnaire, we compared the results obtained for the AMAS online with a previous Polish AMAS study, where data had been collected by paper-and-pencil administration (Cipora et al., 2015a). The comparison is presented in Table 2.

Table 2

Comparison Between AMAS Online and Paper-and-Pencil Administration

Measures		Online study	Paper-and-pencil	Comments
M (*SD*)	Total	20.96 (6.22)	21.93 (6.63)	$t(1470) = -2.85$, $p = .005$, $d = 0.15$; small effect size
	Learning	7.99 (3.10)	8.32 (3.67)	$t(1470) = -1.90$, $p = .057$, $d = 0.10$; small effect size
	Testing	12.97 (3.94)	13.61 (4.01)	$t(1470) = -3.02$, $p = .003$, $d = 0.16$; small effect size
Gender differences in total score		significant $d = 0.34$	significant $d = 0.61$	Higher scores in females, larger gender difference in paper-and-pencil setting
Correlation between scales		.55 (.49-.60)	.49 (.44-.54)	Confidence intervals largely overlap, minimally higher scores in case of online administration
Correlation *Learning* - total		.85 (.83-.87)	.85 (.83-.87)	
Correlation *Testing* - total		.91 (.90-.92)	.88 (.86-.89)	
Reliability – alpha (95% c.i.)	Total	.84 (.82-.86)	.85 (.83-.86)	Confidence intervals largely overlap. lower reliability for *Learning* scale in case of online administration
	Learning	.72 (.68-.75)	.78 (.76-.80)	
	Testing	.85 (.83-.87)	.84 (.82-.86)	
Reliability – ordinal alpha	Total	.88	.88	Virtually identical estimates
	Learning	.80	.84	
	Testing	.88	.87	
Factor structure (exploratory FA, oblique rotation)		58.3% variance explained by two-factor solution	61.6% variance explained by two-factor solution	Factor structure reflects original scales
Confirmatory FA		RMSEA = .079; AGFI = .895	RMSEA = .075; AGFI = .905	The same structural model reaches acceptable fit with observed data

Mean scores differed between administration forms. However, effect sizes related to these differences can be considered as small. In the online study, we observed smaller gender differences than in the paper-and-pencil study.

Correlations between scales and between scales and total scale were very similar and their confidence intervals largely overlap. The same is true for all reliability estimates we used.

Norms

Having administered AMAS in three separate studies (this contribution; Cipora et al., 2015a, 2015b; this study was not published, however, its results were presented as a conference poster which is available at http://doi.org/10.17605/OSF.IO/QB768), we had data from 2057 individuals of varied age and education. Thus, it was possible to establish AMAS norms for different age groups and administration methods. The normative data was prepared for (1) secondary schoolers (grades 7-9); (2) high schoolers (grades 10-12); (3) adults tested with paper-and-pencil questionnaires, and (4) adults tested online. For each group norms were prepared for female and male participants separately. All tables are presented in Appendix B.

Percentile Norms

Percentile norms indicate the percentage of participants who score below and up to a given score (i.e., percentage of participants scoring below + 50% of the participants obtaining a given raw score, according to the proposal by Crawford, Garthwaite, and Slick (2009). To make our norms comparable to those reported by Caviola et al. (2017), in the first step we prepared percentile norms (Table B1).

Standard Norms

Additionally, we prepared standard c-norms (M = 5; SD = 2), since a standard norm is required to be able to compare the performance of an individual participant with inferential statistical procedures (cf. Willmes, 2010). The probability transform was used to approximate normal distributions for originally skewed raw data and smoothing of the raw score distribution was used to mimic a continuous variable. In order to obtain z-quantile scores for each raw score, half of the frequency of that score is subtracted from the cumulative frequency for that score. This percentile rank is taken as the corresponding quantile of a standard normal variable (cf. Gulliksen, 1987). C-scores for adolescents are presented in Table B2 and those for adults, in Table B3.

Standard norms allow for direct comparisons between a participant's scores observed in both scales (Willmes, 2010). Namely, one can check, whether differences in standard scores between scales are unlikely (at a given alpha level) to originate purely from measurement error (so called reliability aspect). Furthermore, it is also possible to check, whether differences in standardized scores are unlikely (at a given alpha level) to occur in the reference population (so called diagnostic validity aspect).

The critical differences for both reliability and validity aspect should be calculated on so called τ-standardized c-scores, which take into consideration possible differences in reliabilities of the scales to be compared regarding differences in true performance level in a given reference group. These scores are reported in Tables B2 and B3 for adolescents and adults respectively.

In Table B4 critical values for differences between τ-standardized c-scores are presented at various alpha levels. In case of diagnostic validity, it seems that a more liberal approach should be adapted, as it may be more important not to commit the type-II error (i.e., overlooking a possible true difference in performance).

Example

Here we present a step-by step example on how to use the norms we provide. Participant A is an adult male and was administered AMAS in the online form. His total raw score was 26, raw scores in the *Learning* and *Testing* scales were 6 and 20 respectively. Thus, based on Table B1, his percentile score corresponding to the total raw score is 80 (i.e., 80% of the reference population do not score higher). Percentile scores for the *Learning* and *Testing* scales are 30 and 95 respectively. His c-scores are 7 for the total score, 5 for the *Learning* scale and an exceptionally high standardized c-score of 11 for the *Testing* scale.

In the next step one might aim to test, whether there is a difference in true standardized scores obtained by participant A in the *Learning* and *Testing* scales or whether this difference of 6 c-scores is unlikely to originate from measurement error only (i.e., the reliability aspect). One needs to first look up τ-standardized c-scores corresponding to participant A's raw scores in Table B3. In case of the *Learning* scale it is 5.0 and for the *Testing* scale it is 11.71. Thus, the difference in τ-standardized c-scores equals 6.71. Subsequently, the obtained difference must be compared with critical values from Table B4. For the reliability aspect, the critical difference at a type-I error level of .05 equals 3.33. The difference observed in Participant A exceeds this value, thus one may say with 95% confidence that the difference does not originate purely from measurement error.

One might additionally ask whether such a difference is unlikely in the reference population (i.e., the diagnostic validity aspect). Again, one needs to compare the difference in the τ-standardized c-scores (i.e., 6.71) with the respective critical value from Table B4. For the validity aspect, the critical difference at 10% in the corresponding group (adult males, online administration) equals 5.94. Thus, in our case one may say that a higher score difference has a probability of less than 10% in the reference population.

Discussion

Overview

In this study, we investigated the usefulness of the AMAS as a tool for online measurement of math anxiety. The results from a large Polish sample provide further support for the validity of the math anxiety construct as well as the quality of the AMAS as a measurement instrument.

Such measurement method invariance across countries and cultures should not be assumed without empirical support. A vast majority of studies on similarities and differences between administration methods available to English-speaking readers were conducted within Anglo-Saxon cultures, including data on AMAS measurement invariance obtained up to date. Middle European countries, despite overall similarities, differ in several aspects from Anglo-Saxon cultures (e.g., Lee, 2009; OECD, 2013). Furthermore, one should also keep in mind that in Middle and Eastern European countries, personal computers and especially affordable fast Internet connections became available for the average user a few years later than in the U.S.

One may argue that general cultural differences between Polish and Anglo-Saxon cultures are rather small and therefore measurement method invariance should not be questioned even if we take into account possible differences in popularity of personal computers. However, some characteristics of math anxiety in Polish adolescents, as shown in the PISA 2012 study (OECD, 2013), point to some specifics that legitimize such studies. Namely, the observed relationship between MA and math performance is one of the strongest across

all countries and economies involved in the PISA program. Furthermore, the gender gap was considerably below the PISA average. Because of these remarkable cultural specificities, it is far from clear that MA results from Western cultures generalize to other cultures, namely, a Slavic culture here. Importantly, the PISA study provides data about adolescents, however much less is known about commonalities and differences in math anxiety in the adult population. However, we observed that the AMAS scale remained invariant across administration methods in Poland as well, with regard to construct validity (reflected by confirmatory factor analysis) and reliability estimates. This is good news for international assessment and cross-cultural research, because it seems that we are assessing largely the same construct, no matter if we do it online or in paper-and-pencil administration.

Online Versus Paper-and-Pencil Measurement

Ashcraft and Faust (1994) claim that higher MA scores observed in case of computerized administration are due to the fact that math anxious individuals do not feel comfortable with computers. We hypothesized that, although this may have been relevant in the early 90's, this claim is no longer valid. Nowadays, computers are much more ubiquitous and user-friendly in general, so most likely they do not induce additional anxiety, especially regarding typical computer usage (e.g., filling out an online survey, which does not differ from typical website activities). A comparison with a previous study (Cipora et al., 2015a) showed a small difference in AMAS scores depending on administration methods. Notably, the difference we observed was in the opposite direction (i.e., participants administered with the online form scored lower rather than higher on MA) than the effect reported by Ashcraft and Faust (1994). Nevertheless, the corresponding effect sizes were very small, and according to Cohen's (1988) recommendations they should not be considered as practically meaningful. Therefore, based on the results of the present study, we conclude that the AMAS can be administered online for practical screening purposes etc. However, if researchers look for small differences between groups (cultures, age groups, gender etc.), they should not administer the AMAS online to one group and per paper-and-pencil in another.

AMAS Structure

The two-factorial structure of the AMAS proposed by its authors was present in Polish participants as well. The only difference is that the item referring to homework can be assigned to both scales. This can make sense, because homework has a learning aspect (because math is practiced at home), but can also have a testing aspect, e.g., when participants anticipate being controlled by a teacher later.

The overall stability of factorial structure can be considered as further evidence of AMAS validity. However, a stable two-factorial structure of the AMAS does not require a two-componential structure of the construct of interest. It is still possible that several aspects of MA may not be covered by AMAS items. For instance, Pletzer et al. (2016) showed that MA consists of more than two components, e.g., anxiety related to being tested and anxiety related to anticipation of an upcoming math test. This is of particular importance when it comes to interventions aimed at reducing aspects of MA (see Pletzer et al., 2016).

However, the AMAS is much shorter than the instrument used by Pletzer et al. (2016). Some potential aspects of math anxiety like social responsibility are not covered in the AMAS. Other aspects may be less covered (e.g., by single items). Therefore, seemingly inconsistent factor structures can be reconciled if one acknowledges that the structure depends on the instrument and the exact questions being asked in it. This discrepancy calls for a

theoretical debate about which items should be included in math anxiety assessment. Currently, it appears that despite its usefulness for screening and research purposes, AMAS usefulness for individual diagnosis, especially in case of differential intervention planning, should be tested in future studies.

Gender Differences in Math Anxiety

Compared to the paper-and-pencil study by Cipora et al. (2015a), gender differences in AMAS scores were considerably smaller. Interestingly, we found that gender differences differed depending on the field of study category and were in fact present only in the STEM group, but not in other fields of studies categories for which we could perform such calculations (i.e., psychology and "neuropsychology, neurobiology, and cognitive science" categories). Importantly, it is also in line with results of a paper-and-pencil study (Cipora et al., 2015a), where no gender differences were observed when only psychology students were considered. However, an individual's selection of a field of study is, among other factors, guided by one's attitudes, interests, stereotypes and math anxiety (Ashcraft & Ridley, 2005). Therefore, this issue requires further attention as far as gender differences in MA are concerned. One must also keep in mind that there may be a selection bias in case of online assessment, possibly causing individuals who are not representative of the general population (and their respective genders) to decide to participate. Unfortunately, in the current study setup, we were not able to obtain data on the number of participants who actually clicked the link but then decided not to submit their answers.

In general, our results suggest that one must be very cautious regarding claims about gender differences in MA. Especially they provide some hints on reasons for inconsistent results as regards gender differences in math anxiety. These inconsistencies may at least to some extent originate from the fact that different samples are tested as regards participants' current field of study. Furthermore, one must also keep in mind that the size of gender differences differs largely between countries and cultures, as also shown in the PISA study.

Results of both exploratory (Appendix A) and confirmatory factor analyses suggest that the factor structure does not differ substantially between genders, which means that apart from gender differences in average math anxiety, the internal structure of the construct remains unchanged across genders.

Limitations of the Presented Study

First of all, the sample of our study was very homogeneous in terms of educational background. Like many previous studies (and psychology studies in general), we tested mostly young adult university students, which limits extrapolating results of our study to the general population. On the other hand, the scores obtained by our participants cover a large part of the AMAS theoretical score range, so that reliability estimates and item characteristics are not affected by low systematic variability in the data.

Nevertheless, it would be worthwhile to test more varied samples, including people with varied highest educational levels reached. In particular, it would be useful to test participants who are not well familiarized with psychological measurement tools in general.

Another important limitation of the presented study was that it did not include any other measure, which precludes the possibility to investigate convergent and discriminant validity. It seems relatively unlikely that two

administration forms being invariant regarding factor structure, reliabilities, mean scores, and variances, differ in correlations with external measures. Nevertheless, this should be addressed in future studies.

Importantly, our study did not involve any math-related activity. The participants were only asked to recall from memory how anxious they will feel if they are or would be confronted with several math-related situations. As thoroughly discussed by Bieg, Goetz, Wolter, and Hall (2015), such trait-like anxiety ratings can differ from those obtained after anxiety-evoking exposure: trait-like ratings are usually overestimates. Therefore, comparing studies in which participants were or were not exposed to math and were assessed for their MA prior or after the exposure cannot be considered equivalent a priori. This also requires future investigation.

General Conclusions

The results of our study provide further evidence for the usefulness of the AMAS as a math anxiety measure. Cross–cultural invariance, together with measurement method invariance, strongly suggests validity and generalizability of the MA construct. Interestingly, average AMAS scores reported from different countries also seem to be relatively similar (Dykeman, 2017). Furthermore, the combined results of numerous studies suggest that the AMAS validly reflects the MA construct. Further research is still needed to examine the fine-grained structure of MA. It also seems that the AMAS may be useful as a MA measurement instrument in varied settings, i.e., paper-and-pencil, computerized, and online. Therefore, the AMAS may be useful as an additional measure in various fields of numerical cognition since there is a growing body of evidence for the role of MA to be involved in several aspects of human number processing. In order to make practical use of what we know on math anxiety, it is important that practitioners are equipped with normative data so that they may make more informed choices about individuals who are potentially in threat of high math anxiety. We hope that future researchers will also make an extra effort to prepare adequate AMAS norms for several populations.

Supplementary Materials

Raw data and analytic code can be found at https://doi.org/10.17605/OSF.IO/Y76FS.

Supplementary material comprises raw data file (txt file, tab separated) as well as R script which allows reproducing all analyses reported in the paper as well as accessing some extra results. Some parts of the analyses (i.e., estimating *SE* for skewness and kurtosis as well as exploratory factor analyses were conducted with IBM SPSS 24. Shared SPSS syntax allows reproducing these analyses as well.

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

References

Alexander, L., & Martray, C. R. (1989). The development of an abbreviated version of the Mathematics Anxiety Rating Scale. *Measurement & Evaluation in Counseling & Development, 22*(3), 143-150.

Arbuckle, J. L. (2016). *Amos* (Version 24.0) [Computer program]. Chicago, IL, USA: IBM SPSS.

Artemenko, C., Daroczy, G., & Nuerk, H.-C. (2015). Neural correlates of math anxiety – An overview and implications. *Frontiers in Psychology, 6*, Article 1333. doi:10.3389/fpsyg.2015.01333

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

Ashcraft, M. H., & Faust, M. W. (1994). Mathematics anxiety and mental arithmetic performance: An exploratory investigation. *Cognition and Emotion, 8*(2), 97-125. doi:10.1080/02699939408408931

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* (p. 315-327). New York, NY, USA: Psychology Press.

Bai, H. (2011). Cross-validating a bidimensional mathematics anxiety scale. *Assessment, 18*(1), 115-122. doi:10.1177/1073191110364312

Baloğlu, M. (2003). Individual differences in statistics anxiety among college students. *Personality and Individual Differences, 34*(5), 855-865. doi:10.1016/S0191-8869(02)00076-4

Bedyńska, S., & Książek, M. (2012). *Statystyczny drogowskaz 3. Praktyczny przewodnik wykorzystania modeli regresji oraz równań strukturalnych.* Warsaw, Poland: Wydawnictwo Akademickie Sedno.

Bieg, M., Goetz, T., Wolter, I., & Hall, N. C. (2015). Gender stereotype endorsement differentially predicts girls' and boys' trait-state discrepancy in math anxiety. *Frontiers in Psychology, 6*, Article 1404. doi:10.3389/fpsyg.2015.01404

Booth-Kewley, S., Edwards, J. E., & Rosenfeld, P. (1992). Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? *The Journal of Applied Psychology, 77*(4), 562-566. doi:10.1037/0021-9010.77.4.562

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. doi:10.5430/jct.v5n2p76

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science, 332*(6033), 1049-1053. doi:10.1126/science.1201536

Carey, E., Hill, F., Devine, A., & Szűcs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology, 6*, Article 1987. doi:10.3389/fpsyg.2015.01987

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*, Article 11. doi:10.3389/fpsyg.2017.00011

Carey, E., Devine, A., Hill, F., & Szűcs, D. (2017b). Differentiating anxiety forms and their role in academic performance from primary to secondary school. *PLoS One, 12*(3), Article e0174418. doi:10.1371/journal.pone.0174418

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. doi:10.1016/j.lindif.2017.03.006

Cipora, K., Szczygieł, M., Willmes, K., & Nuerk, H.-C. (2015a). Math anxiety assessment with the Abbreviated Math Anxiety Scale: Applicability and usefulness: Insights from the Polish adaptation. *Frontiers in Psychology, 6*, Article 1833. doi:10.3389/fpsyg.2015.01833

Cipora, K., Szwarc, A., Zaremba, D., Konieczny, J., Machniak, A., Patrzałek, O., . . . Nuerk, H.-C. (2015b). *Math anxiety evaluation in Polish adolescents with Abbreviated Math Anxiety Scale (AMAS).* Poster session presented at the workshop "Neuroeducation of Number Processing", Hannover, Germany. doi:10.17605/OSF.IO/QB768

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

Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009). On percentile norms in neuropsychology: Proposed reporting standards and methods for quantifying the uncertainty over the percentile ranks of test scores. *The Clinical Neuropsychologist, 23*, 1173-1195. doi:10.1080/13854040902795018

Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions, 8*, Article 33. doi:10.1186/1744-9081-8-33

Dietrich, J. F., Huber, S., Moeller, K., & Klein, E. (2015). The influence of math anxiety on symbolic and non-symbolic magnitude processing. *Frontiers in Psychology, 6*, Article 1621. doi:10.3389/fpsyg.2015.01621

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

Dykeman, C. (2017, January). The weighted average of Abbreviated Math Anxiety Scale (AMAS) studies on College Students. *Technical Bulletin, 2017-01*. Retrieved from http://ir.library.oregonstate.edu/xmlui/bitstream/handle/1957/60103/AMAS%20Technical%20Bulletin%20Dykeman%201-9-17.pdf

Ferguson, A. M., Maloney, E. A., Fugelsang, J., & Risko, E. F. (2015). On the relation between math and spatial ability: The case of math anxiety. *Learning and Individual Differences, 39*, 1-12. doi:10.1016/j.lindif.2015.02.007

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17*(3), 1-13. Retrieved from http://pareonline.net/getvn.asp?v=17&n=3

Gulliksen, H. (1987). *Theory of mental tests*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Gwaltney, C. J., Shields, A. L., & Shiffman, S. (2008). Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value in Health, 11*(2), 322-333. doi:10.1111/j.1524-4733.2007.00231.x

Gyuris, E., & Everingham, Y. (2011, August). Maths anxiety over two campuses in a first year introductory quantitative skills subject at a regional Australian university – Establishing a baseline. In *Proceedings of The Australian Conference on Science and Mathematics Education (formerly UniServe Science Conference)* (Vol. 17, pp. 73-80). Retrieved from https://openjournals.library.sydney.edu.au/index.php/IISME/article/view/4831

Gyuris, E., Everingham, Y., & Sexton, J. (2012). Maths anxiety in a first year introductory quantitative skills subject at a regional Australian university – Establishing a baseline. *International Journal of Innovation in Science and Mathematics Education, 20*, 42-54. Retrieved from https://openjournals.library.sydney.edu.au/index.php/CAL/article/view/6647

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

Hopko, D. R. (2003). Confirmatory factor analysis of the Math Anxiety Rating Scale–Revised. *Educational and Psychological Measurement, 63*(2), 336-351. doi:10.1177/0013164402251041

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. doi:10.1177/1073191103010002008

Hunt, T. E., Clark-Carter, D., & Sheffield, D. (2011). The development and part validation of a UK scale for mathematics anxiety. *Journal of Psychoeducational Assessment, 29*(5), 455-466. doi:10.1177/0734282910392892

Jones, W. J., Childers, T. L., & Jiang, Y. (2012). The shopping brain: Math anxiety modulates brain responses to buying decisions. *Biological Psychology, 89*(1), 201-213. doi:10.1016/j.biopsycho.2011.10.011

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Chicago, IL, USA: Scientific Software International Inc.

Kazelskis, R. (1998). Some dimensions of mathematics anxiety: A factor analysis across instruments. *Educational and Psychological Measurement, 58*(4), 623-633. doi:10.1177/0013164498058004006

King, W. C., & Miles, E. W. (1995). A quasi-experimental assessment of the effect of computerizing noncognitive paper-and-pencil measurements: A test of measurement equivalence. *Journal of Applied Psychology, 80*(6), 643-651. doi:10.1037/0021-9010.80.6.643

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences, 19*(3), 355-365. doi:10.1016/j.lindif.2008.10.009

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*(5), 520-540. doi:10.2307/749772

Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. *Journal of Adolescence, 27*(2), 165-179. doi:10.1016/j.adolescence.2003.11.003

Núñez-Peña, M. I., Suárez-Pellicioni, M., Guilera, G., & Mercadé-Carranza, C. (2013). A Spanish version of the short Mathematics Anxiety Rating Scale (sMARS). *Learning and Individual Differences, 24*, 204-210. doi:10.1016/j.lindif.2012.12.009

OECD. (2010). *The high cost of low educational performance: The long-run economic impact of improving educational outcomes*. Paris, France: OECD.

OECD. (2013). *PISA 2012 results: Ready to learn: Students' engagement, drive and self-beliefs* (Vol. 3). Paris, France: OECD Publishing.

Pettit, F. A. (2002). A comparison of World-Wide Web and paper-and-pencil personality questionnaires. *Behavior Research Methods, Instruments, & Computers, 34*(1), 50-54. doi:10.3758/BF03195423

Pletzer, B., Kronbichler, M., Nuerk, H.-C., & Kerschbaum, H. H. (2015). Mathematics anxiety reduces default mode network deactivation in response to numerical tasks. *Frontiers in Human Neuroscience, 9*, Article 202. doi:10.3389/fnhum.2015.00202

Pletzer, B., Wood, G., Scherndl, T., Kerschbaum, H. H., & Nuerk, H.-C. (2016). Components of mathematics anxiety: Factor modeling of the MARS30-Brief. *Frontiers in Psychology, 7*, Article 91. doi:10.3389/fpsyg.2016.00091

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. doi:10.1016/j.lindif.2014.05.012

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org

Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology, 49*(4), 243-256. doi:10.1026//1618-3169.49.4.243

Reips, U.-D., Buchanan, T., Krantz, J., & McGraw, K. (2015). Methodological challenges in the use of the Internet for scientific research: Ten solutions and recommendations. *Studia Psychologica, 15*, 139-148. doi:10.21697/sp.2015.14.2.09

Ritter, P., Lorig, K., Laurent, D., & Matthews, K. (2004). Internet versus mailed questionnaires: A randomized comparison. *Journal of Medical Internet Research, 6*(3), Article e29. doi:10.2196/jmir.6.3.e29

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74. Retrieved from http://www.dgps.de/fachgruppen/methoden/mpr-online

Sheng, Y., & Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? *Frontiers in Psychology, 3*, Article 34. doi:10.3389/fpsyg.2012.00034

Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, À. (2016). Math anxiety: A review of its cognitive consequences, psychophysiological correlates, and brain bases. *Cognitive, Affective & Behavioral Neuroscience, 16*(1), 3-22. doi:10.3758/s13415-015-0370-7

Suinn, R. M., & Edwards, R. (1982). The measurement of mathematics anxiety: The Mathematics Anxiety Rating Scale for Adolescents – MARS-A. *Journal of Clinical Psychology, 38*(3), 576-580. doi:10.1002/1097-4679(198207)38:3<576::AID-JCLP2270380317>3.0.CO;2-V

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.

Wang, Z., Hart, S. A., Kovas, Y., Lukowski, S., Soden, B., Thompson, L. A., . . . Petrill, S. A. (2014). Who is afraid of math? Two sources of genetic variance for mathematical anxiety. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 55*(9), 1056-1064. doi:10.1111/jcpp.12224

Willmes, K. (2010). The methodological and statistical foundations of neuropsychological assessment. In J. Gurd, U. Kischka, & J. Marshall (Eds.), *The handbook of clinical neuropsychology* (2nd ed., p. 28-49). New York, NY, USA: Oxford University Press.

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment, 29*(4), 377-392. doi:10.1177/0734282911406668

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*(1), 21-29. doi:10.22237/jmasm/1177992180

Appendices

Appendix A: Exploratory Factor Analysis

In addition to confirmatory factor analysis, we also conducted an exploratory factor analysis on the AMAS online results. The analysis was carried out for all participants and for females and males separately. The data were suitable for factor analysis (KMOs =.86, .87 and .79 for the whole sample, females, and male participants, respectively). The Bartlett test for deviation from sphericity was significant at *p* < .001 in all three instances. To make the results comparable with analogous analyses conducted on data from paper-and-pencil testing and to fit it to our data, we used the principal axes factor (PFA) extraction method. This method is more suitable than principal component analysis in cases when multivariate normality is violated. From a theoretical point of view, it is possible that the factors are correlated, therefore we used an oblique rotation (Simple OBLIMIN). Based on the scree plot, we found two factors accounting for 58.34% of the total variance in the data. The first component (corresponding to the *Testing* scale) accounted for 44.61%, and the second (corresponding to the *Learning* scale) for additional 13.73%. Factor loadings (model matrix) for the items are presented in Table A1.

For female participants, two factors accounted for 59.89% of the total variance. The first, corresponding to *Testing,* explained 46.13% of variance and the second, corresponding to *Learning*, explained an additional 13.75%. For male participants, 2 factors were found that accounted for 54.66% of variance. The first factor, corresponding to *Testing*, explained 40.11% of variance, and the second, corresponding to *Learning*, explained an additional 14.55% of variance.

In general, the factor structure (2-factorial solution and proportion of variance explained) is very similar to the paper-and-pencil data. However, in the case of paper-and-pencil data, the first factor corresponded to the *Learning* scale, while in the online data, it corresponded to the *Testing* scale. This difference seems to originate from mathematical properties of the oblique rotation. When an orthogonal (VARIMAX) rotation (PFA extraction method) was used, the first factor corresponded to the *Testing* scale in both instances. In general, results of exploratory factor analysis also correspond well with the structure of the original English version of the AMAS.

Table A1

Factor Loadings of AMAS Items Estimated by Means of Exploratory Factor Analysis (Simple OBLIMIN Rotation)

		EFA factor loadings					
		All participants (*n* = 615)		Females (*n* = 418)		Males (*n* = 197)	
Item	Scale	1	2	1	2	1	2
1	*Learning*	.15	.32	.12	.32	.22	.35
2	*Testing*	.82	.06	.80	.09	.88	-.04
3	*Learning*	.09	.64	.06	.67	.21	.57
4	*Testing*	.81	-.06	.81	-.05	.77	-.06
5	*Testing*	.47	.31	.51	.27	.37	.40
6	*Learning*	-.12	.78	-.10	.77	-.15	.79
7	*Learning*	-.04	.60	-.05	.66	-.03	.46
8	*Testing*	.81	-.02	.88	-.05	.66	.04
9	*Learning*	.23	.40	.19	.45	.29	.30

Appendix B: AMAS Norms

Table B1

Percentile Norms of the AMAS Scale.

Scale	Cronbach's alpha	*M*	*SD*	Percentile										
				5	10	20	30	40	50	60	70	80	90	95
Secondary schoolers, females (*n* = 141). Form: paper														
total	.82	18.12	6.15	9	11	12	15	16	17	19	21	25	27	29
Learning	.72	7.28	2.95	5	5	5	5	5	6	7	8	10	12	13
Testing	.83	10.84	4.11	4	6	7	8	9	11	12	13	14	17	18
Secondary schoolers, males (*n* = 119). Form: paper														
total	.80	17.76	6.47	9	10	12	14	15	17	18	20	23	28	29
Learning	.77	7.71	3.70	5	5	5	5	6	6	7	8	9	13	17
Testing	.83	10.06	4.29	4	4	6	7	8	9	11	12	14	17	18
High schoolers, females (*n* = 213). Form: paper														
total	.82	20.82	6.08	11	13	16	17	19	20	22	23	26	29	32
Learning	.74	7.79	3.26	5	5	5	6	6	7	7	9	10	13	15
Testing	.81	13.02	3.84	6	7	10	11	12	13	14	15	17	18	19
High schoolers, males (*n* = 111). Form: paper														
total	.86	18.97	7.05	9	11	13	14	16	19	20	22	24	29	31
Learning	.83	7.87	3.69	5	5	5	5	6	7	7	9	10	13	15
Testing	.85	11.10	4.36	4	5	7	8	10	11	13	13	15	17	19
Adults, females (*n* = 688). Form: paper														
total	.85	22.63	6.63	13	14	17	19	21	22	24	26	28	32	35
Learning	.79	8.47	3.74	5	5	5	6	7	7	8	9	11	14	16
Testing	.83	14.16	3.85	7	9	11	12	13	15	16	17	18	19	20
Adults, males (*n* = 160). Form: paper														
total	.77	18.86	5.68	10	12	14	16	17	18	19	21	23	26	30
Learning	.69	7.59	3.15	5	5	5	5	6	7	7	8	10	12	14
Testing	.81	11.27	3.85	4	6	8	9	10	12	12	13	15	16	18
Adults, females (*n* = 418). Form: online														
total	.85	21.62	6.33	13	14	16	17	20	21	23	25	27	30	33
Learning	.74	8.17	3.22	5	5	5	6	7	7	8	9	11	13	15
Testing	.87	13.45	3.95	7	8	10	11	13	14	15	16	17	18	19
Adults, males (*n* = 197). Form: online														
total	.81	19.54	5.73	11	13	14	16	18	18	21	23	25	28	30
Learning	.68	7.59	2.80	5	5	5	5	6	7	8	9	10	12	13
Testing	.80	11.95	3.73	6	7	9	10	11	12	13	14	16	17	18

Table B2

Standard Norms for Adolescents

	AMAS Total				AMAS Learning								AMAS Testing							
	Secondary schoolers		High schoolers		Secondary schoolers				High schoolers				Secondary schoolers				High schoolers			
	females	males	females	males	females		males		females		males		females		males		females		males	
c-score	Raw	Raw	Raw	Raw	r-std c	Raw	r-std c	Raw	r-std c	Raw	r-std c	Raw	r-std c	Raw	r-std c	Raw	r-std c	Raw	r-std c	Raw
-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-1.67	4	-	-
0	-	-	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	9	-	11	9	-	-	-	-	-	-	-	-	0.61	4	-	-	0.56	5	0.66	4
2	10	9-10	12-13	10	-	-	-	-	-	-	-	-	1.71	5	1.71	4	1.67	6-7	1.75	5
3	11-12	11-12	14-16	11-13	2.64	5	2.72	5	2.68	5	2.80	5	2.80	6-7	2.80	5-6	2.78	8-10	2.83	6-7
4	13-15	13-15	17-18	14-16	-	-	-	-	3.84	6	-	-	3.90	8-9	3.90	7-8	3.89	11-12	3.92	8-9
5	16-18	16-18	19-21	17-19	5.00	6	5.00	6-7	5.00	7	5.00	6-7	5.00	10-11	5.00	9-10	5.00	13-14	5.00	10-12
6	19-23	19-21	22-24	20-22	6.18	7-9	6.14	8	6.16	8-9	6.10	8-9	6.10	12-14	6.10	11-13	6.11	15-16	6.08	13-14
7	24-26	22-27	25-28	23-28	7.36	10-11	7.28	9-12	7.32	10-12	7.20	10-12	7.20	15-16	7.20	14-16	7.22	17	7.17	15-16
8	27-29	28-31	29-33	29-31	8.54	12-14	8.42	13-17	8.49	13-15	8.29	13-17	8.29	17-18	8.29	17-18	8.33	18-19	8.25	17-19
9	30-32	32-36	34-36	32-44	9.71	15-17	9.56	18-25	9.65	16	9.39	18-24	9.39	19	9.39	19	-	-	9.34	20
10	33-45	37-45	37-42	45	10.89	18-25	-	-	10.81	17-22	10.49	25	10.49	20	10.49	20	10.56	20	-	-
11	-	-	43-45	-	-	-	-	-	11.97	23-25	-	-	-	-	-	-	-	-	-	-

Table B3

Standard Norms for Adults.

c-score	AMAS Total				AMAS Learning								AMAS Testing							
	Paper Administration		Online administration		Paper Administration				Online administration				Paper Administration				Online administration			
	females	males	females	males	females		males		females		males		females		males		females		males	
					t-std c	Raw	t-std c	Raw	t-std c	Raw	t-std c	Raw	t-std c	Raw	t-std c	Raw	t-std c	Raw	t-std c	Raw
-1	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0	10	-	9-10	-	-	-	-	-	-	-	-	-	-	-	-	-	-0.36	4	-	-
1	11-12	9	11-12	9-10	-	-	-	-	-	-	-	-	0.61	4-5	0.56	4	0.71	5-6	0.53	4-5
2	13-14	10-12	13-14	11-12	-	-	-	-	-	-	-	-	1.71	6	1.67	5-6	1.78	7-8	1.65	6-7
3	15-17	13-14	15-16	13-14	2.75	5	2.59	5	2.68	5	2.57	5	2.80	7-8	2.78	7-8	2.86	9	2.76	8
4	18-20	15-17	17-19	15-17	3.87	6	-	-	3.84	6	-	-	3.90	9-11	3.89	9-10	3.93	10-12	3.88	9-10
5	21-23	18-19	20-22	18-20	5.00	7-8	5.00	6-7	5.00	7-8	5.00	6-7	5.00	12-13	5.00	11-12	5.00	13-14	5.00	11-12
6	24-27	20-22	23-26	21-24	6.13	9-10	6.20	8-9	6.16	9-10	6.21	8-9	6.10	14-15	6.11	13-14	6.07	15-16	6.12	13-15
7	28-31	23-25	27-30	25-27	7.25	11-13	7.41	10-12	7.32	11-12	7.43	10-11	7.20	16-17	7.22	15-16	7.14	17-18	7.24	16
8	32-35	26-30	31-33	28-30	8.38	14-16	8.61	13-14	8.49	13-15	8.64	12-13	8.29	18	8.33	17-18	8.22	19	8.35	17-18
9	36-40	31-34	34-36	31-33	9.50	17-21	9.82	15-18	9.65	16-17	9.85	14-15	9.39	19	9.44	19	9.29	20	9.47	19
10	41-43	35-45	37-44	34-45	10.63	22-23	11.02	19-25	10.81	18-24	-	-	10.49	20	10.56	20	-	-	-	-
11	44-45	-	45	-	11.75	24-25	-	-	11.97	25	12.28	16-25	-	-	-	-	-	-	11.71	20

Table B4

Critical Differences Between Scores at Different Alpha Levels

Gender	Between scale correlation	SEM 1 (reliability aspect)				SEM 2 (validity aspect)				
			Critical difference (in τ-std c-scores)				Critical difference (in τ-std c-scores)			
		SEM 1	1%	5%	10%	SEM 2	1%	5%	10%	20%
Secondary schoolers, Form: paper										
females	0.50	1.54	3.97	3.02	2.53	3.43	8.85	6.72	5.64	4.42
males	0.31	1.42	3.66	2.78	2.34	2.99	7.71	5.86	4.92	3.86
High schoolers. Form: paper										
females	0.46	1.53	3.95	3.00	2.52	3.36	8.67	6.59	5.53	4.33
males	0.53	1.23	3.17	2.41	2.02	3.28	8.46	6.43	5.40	4.23
Adults. Form: paper										
females	0.53	1.37	3.53	2.69	2.25	3.37	8.69	6.61	5.54	4.35
males	0.31	1.65	4.26	3.23	2.71	3.15	8.13	6.17	5.18	4.06
Adults. Form: online										
females	0.56	1.42	3.66	2.78	2.34	3.44	8.88	6.74	5.66	4.44
males	0.53	1.70	4.39	3.33	2.80	3.61	9.31	7.08	5.94	4.66