Original Article

Not Only Reliability!

The Importance of the Ecological Validity of the Math Anxiety Questionnaire for Adults

Monika Szczygieł

Institute of Psychology, Pedagogical University of Krakow, Poland

Abstract: The measurement of math anxiety in adults is justified based on observations that math anxiety in parents and teachers predicts children's math anxiety and achievement. Although there are many very good math anxiety measures intended for children and adolescents, their usefulness (e.g., AMAS, MARS) for adults is debatable. The most important objection against using these scales for adults is their ecological validity. The measurement of anxiety associated with math tests, classes, teachers, and homework is adequate for students of science, technology, engineering, and mathematics (STEM), but not for students of social sciences and humanities (HS) and non-students (e.g., parents and preschool and early education teachers). In response to this gap, the Math Anxiety Questionnaire for Adults (MAQA) was developed; it is designed to measure math anxiety related to math problem-solving in various groups of adults (especially non-students and HS students, as well as STEM students). The content, construct, criterion, and ecological validity of the MAQA were tested, and its internal and test-retest reliability was established. The results confirm that the MAQA is a valid and reliable measurement of math anxiety; therefore, it may be recommended for use in various groups of adults (e.g., students, teachers, and parents).

Keywords: math anxiety, measurement, validity, reliability, adults

OPEN DATA OPEN MATERIALS

Math Anxiety Definition, Dimensions, and Measurement in Adults

Math anxiety is typically defined as "[...] a feeling of tension and anxiety that interferes with the manipulation of numbers and the solving of the mathematical problems in a wide variety of ordinary life and academic situations" (Richardson & Suinn, 1972, p. 551). Math anxiety is a multidimensional construct and its various dimensions are tested: (1) in adults and adolescents (math learning and math testing anxiety: Hopko et al., 2003; math evaluation anxiety, everyday/social math anxiety, math observation anxiety: Hunt et al., 2011; numerical anxiety, math test anxiety: Richardson & Suinn, 1972); and (2) children (negative reactions, numerical confidence, worry: Harari et al., 2013; numerical processing anxiety, situational and performance anxiety: Wu et al., 2012; math evaluation and math testing anxiety: Carey et al., 2017). Although it seems that the measurement of everyday math anxiety is more ecologically valid for adults, and the measurement of academic math anxiety is more valid for school learners, similar scales are used for the evaluation of math anxiety in children, adolescents, and adults.

The most popular math anxiety scales are the Mathematics Anxiety Rating Scale (MARS; Richardson & Suinn, 1972), the revised MARS (MARS-R; Plake & Parker, 1982), the shortened MARS (sMARS; Alexander & Martray, 1989), the Abbreviated Math Anxiety Scale (AMAS; Hopko et al., 2003), the Single-Item Math Anxiety Scale (SIMA; Núñez-Peña et al., 2014), and the Math Anxiety Scale-UK (MAS-UK; Hunt et al., 2011). The first four scales were originally intended for testing various types of math anxiety in children or adolescents; however, they were also expanded to be used in adults (Cipora et al., 2015; Hart & Ganley, 2019; Vahedi & Farrokhi, 2011; Hadley & Dorward, 2011; Hopko et al., 2003; Maloney et al., 2015; Soni & Kumari, 2017). The last of the scales are intended for adults and contains both school-related and non-academic subscales (Hunt et al., 2011).

The rationale for using these scales for adults is mainly their high reliability; however, reliability is only one indicator of a good scale, the other being its validity. Newton and Shaw (2019) highlighted that many types of validity are discussed in psychometric literature, but it is constructed, content, and criterion validity that is among the most basic and

https://doi.org/10.1027/1015-5759/a000646

tested types of validity that are mentioned in psychological studies. All these types of relevance should be put under the umbrella term of ecological validity (Schmuckler, 2001). Unfortunately, it seems that measurements of math anxiety used for adults have not fully met this recommendation.

The adult population is diverse in their commitment to learning mathematics in their daily lives. All adults deal with math every day, for example, while making calculations in a store, planning renovation, settling taxes, answering mathematical questions from their own children, and solving crosswords and quizzes, but only some adults have the experience of formal math education in adulthood. A group of adults who still participate in math classes are students. Students of science, technology, engineering, and mathematics (STEM) learn subjects that require mathematical knowledge in many courses. However, students of humanities and social sciences (HS) mostly do not have mathematical courses; the math-related subjects they study at college are most often logic and statistics. In conclusion, probably only those adults who study or practice a STEM profession participate in formal math education, which in turn brings them closer to the experience of a typical learner in a math classroom context.

When asked about the level of math anxiety related to math exams, math pop-quizzes, or math homework STEM students probably think these questions make sense. However, students of literature, art, or pedagogy may find these questions pointless because they do not have math exams or math pop-quizzes, and they do not solve math homework. Since HS students have recently completed their formal math education, they may conclude that these questions relate to their recent past experiences. However, non-students, who for example finished their education 15 years ago, and work, as cleaners, bus drivers, or managers, may be confused and surprised when asked questions about their present anxiety related to math teachers, math classes, or math homework.

When faced with academic-related items (e.g., the MARS, the AMAS, some items in the MAS-UK), non-STEM adults may answer math anxiety questions based on their previous school experiences rather than current ones. Therefore, their score in the questionnaire may not reflect the current level of their math anxiety. Adults may adopt different strategies while answering questions that do not fit their current situation, but the researcher does not know what these strategies are. This does not mean, however, that math anxiety in adults is independent of the school context. On the contrary, the level of math anxiety in adulthood is determined by previous school experiences. Nevertheless, the questions I found ecologically invalid concern math learning, math testing, math teachers, and math classes. Math anxiety in adults may be manifested with a negative reaction toward everyday calculations or helping others with math problems, but not in a feeling of tension related to a math exam that they will never take in the future.

The examination of math anxiety in adults is important for future generations for educational, social, and health reasons (Casad et al., 2015; Hadley & Dorward, 2011; Maloney et al., 2015; Soni & Kumari, 2017). Math anxiety appears and develops from early childhood, and it is believed that the level of math anxiety and mathematical achievement in learners may depend on the level of math anxiety in early education teachers and parents (Beilock et al., 2010; Casad et al., 2015; Maloney et al., 2015; Ramirez et al., 2018; Soni & Kumari, 2017; Szczygieł, 2020a). Early education teachers and parents have experience in solving math problems because they provide children with mathematical knowledge and help children to learn. To date, many interesting scales for studying math teaching anxiety in teachers have been developed (Ganley et al., 2019; Hunt & Sari, 2019), but they cannot be used for investigating parental math anxiety. Some questions/subscales from scales intended for adults could be used for the purpose (e.g., MAS-UK), but in practice, researchers often use scales in their entirety. Thus, a separate scale intended to measure anxiety-related to math problem solving was designed especially with a view to studying the level of math anxiety in parents of learners, preschool and early education teachers, and teacher candidates.

Objectives and Hypotheses of the Present Study

The main objective of the study was to present the psychometric properties of the Math Anxiety Questionnaire for Adults (MAQA). It is a new non-school-dependent questionnaire that does not contain items related to formal math education. It was developed to measure the level of anxiety related to math problem-solving in a wide variety of ordinary and academic life situations. The first aim of the MAQA was to avoid questions about past experiences related to math teachers, math classes, or math exams; the second objective was to check the current level of math anxiety in adults when in contact with various simple mathematical problems. The theoretical assumption underlying the preparation of the scale was that math anxiety manifests among adults when they are faced with calculations and when they solve math problems (Richardson & Suinn, 1972).

It was assumed that the scale would be unidimensional and it would represent anxiety related to solving various mathematical problems. Moreover, it was assumed that there should be no gender or group differences (non-students, STEM students, and HS students) in the structure and properties of the scale in three dimensions: configural, metric, and scalar (Lubiewska & Głogowska, 2018). When determining construct validity, it was assumed that women would have a higher level of math anxiety than men (Ashcraft, 2019; Devine et al., 2012). Furthermore, it was supposed that STEM students would have the lowest math anxiety in comparison to HS students and non-students and that HS students would have a higher level of math anxiety than non-students. This assumption was due to the fact that among non-students there might be people with different education and occupation levels (including STEM and HS areas).

It was assumed that the convergent validity of the scale would be proper if the MAQA score highly and positively correlated with the scores of other math anxiety scales (Hopko et al., 2003). This assumption was due to the fact that various dimensions of math anxiety are closely related to each other. Therefore, it was expected that math problem-solving anxiety should yield comparable correlations with general math anxiety (SIMA) and school-context total math anxiety (AMAS). Moreover, if the MAQA was convergently valid, its score should be negatively and highly related to total result in math attitudes (MASA, Durrani & Tariq, 2009), and moderately/strongly related to MASA dimensions (stronger for the affective dimension of attitude in comparison to cognitive and behavioral dimensions because this subscale also contains feelings of anxiety). Moreover, the MAQA should be moderately/strongly related to math self-concept (MSC, Eden et al., 2013), and math self-efficacy (MSE, Eden et al., 2013). Finally, the MAQA results should be moderately and negatively correlated with math achievement (grades and the high school exit exam; Zhang et al., 2019). These assumptions were due to the fact that with the increase in math anxiety, one could expect an increase in negative emotions in relation to mathematics (the affective dimension of math attitude, MASA-A), an increase in negative beliefs about mathematics (the cognitive dimension of math attitude, MASA-C), and a decrease in engagement when solving mathematical problems (the behavioral dimension of math attitude, MASA-B). Additionally, if anxiety related to math problem solving increased, the level of math self-concept and math self-efficacy should decrease. Finally, if the MAQA score was higher, the level of math grades and math exam scores should be lower. Divergent validity could be confirmed if the MAQA result was positively and weakly correlated with state and trait anxiety (STAI X-1, X-2, Hopko et al., 2003). This is because math anxiety is specific to mathematics but not completely independent from trait and state anxiety (Hart & Ganley, 2019). The divergent validity of the MAQA should also be confirmed if Polish self-concept (PSC), self-efficacy (PSE), and achievement (grades and the high school exit exam [Matura]) did not relate or were very weakly correlated with the MAQA (Cipora et al., 2015).

Finally, it was assumed that the MAQA score and other math anxiety scales (the SIMA and the AMAS) should yield comparable correlations with state and trait anxiety, math attitude, math self-efficacy, and math achievement, as well as Polish self-concept, Polish self-efficacy, and Polish achievement. If these correlation patterns were similar, it would mean that the MAQA scale was equally good as the SIMA and the AMAS in terms of criterion validity. Furthermore, the relationships between the MAQA and the SIMA (1) and the MAQA and the AMAS (2) should be slightly weaker than the correlations between the SIMA and the AMAS assuming that both the SIMA and the AMAS measure more school-related math anxiety than the MAQA.

Method

Participants

Altogether 1,237 adults (886 women, 331 men, 20 no data) participated in the validation study (age: M = 24.48, SD = 6.98, 18–55 years, N = 1,029, not all participants reported their age). The examined group consisted of non-students (M = 39.04, SD = 4.94, range = 28–55, N [women/men] = 364 [294/70]), STEM students (M = 22.10, SD = 2.21, range = 18–41, N [women/men] = 424 [221/202]), and HS students (M = 21.52, SD = 2.97, range = 18–45, N [women/men] = 449 [371/59]). Participation in the study was voluntary and the participants were not rewarded. The sample size was determined taking into account the minimum number of observations needed to perform planned statistical analyzes (Kyriazos, 2018; Martin & Hollins Martin, 2017).

Materials

State and Trait Anxiety

The State-Trait Anxiety Inventory (STAI; the Polish version of the scale was used, Spielberger et al., 2011) consists of two subscales, one of which (X-1) measures state anxiety (e.g., "I feel nervous") and the other (X-2) measures trait anxiety (e.g., "I worry too much over something that really doesn't matter"). Each subscale consists of 20 items with answers on a 4-point scale (1 = *definitely not*, 2 = *rather not*, 3 = *rather yes*, 4 = *definitely yes*). A higher score means a higher level of anxiety. The reliability of X-1 was α = .94 and X-2 α = .89.

Math Anxiety

The Math Anxiety Questionnaire for Adults (MAQA) was developed in accordance with two intentions: to avoid questions about experiences related to math teachers, math classes, or math exams, and to check the current level of adults' math anxiety related to math problem-solving. The MAQA was prepared in several stages.

In the first phase of the questionnaire construction, 34 items were proposed based on Richardson and Suinn's (1972) definition of math anxiety as anxiety and tension related to the manipulation of numbers and the solving of mathematical problems in ordinary life and academic situations. Therefore, the items concerned specific mathematical problems relating to very basic mathematical knowledge (e.g., giving an example of a sequence of natural numbers, determining the probability that it is Wednesday today) and referring to situations that may take place in everyday life (e.g., comparing offers, interest rates on bank deposits, the calculation of average fuel consumption in a car). The first version of the instructions and the proposal of the response scale were also prepared. All the items were evaluated by seven judges (psychologists) for their face, content, and construct validity. The items were also proofread. The response scale and the instructions were also evaluated and discussed. The 4-point response scale was chosen as best. After initial analysis, 25 items were selected for pilot studies. The instruction was as follows:

> "Below there is a list of problems related to mathematics. Imagine that you are in a situation in which it is necessary to solve the following problems and determine if you are anxious in these situations by marking the answers as follows: 1 = *I definitely do not feel anxiety*, 2 = *I rather do not feel anxiety*, 3 = *I rather feel anxiety*, 4 = *I definitely feel anxiety*."

The higher the score in the MAQA, the higher the level of math anxiety.

Three pilot studies were conducted (N = 95, N = 205, N = 59) in the STEM and the HS student samples. In the first and the second study, information about the internal consistency and the test-retest and convergent and divergent validity of the scale was collected. Moreover, qualitative information on the content of the items, the instructions, and the response scale was also collected. The pilot studies indicated that the scale was unidimensional, and it had high reliability and satisfactory validity. In both studies, all items showed satisfactory psychometric properties, but it was decided to remove six items following feedback on their meaning from the respondents. Therefore, the MAQA tested in the final pilot session contained 19 items. It was tested whether the deletion of the items from the MAQA lowered the reliability of the scale, but this did not happen. All the data collected in the pilot sessions suggested that the scale was appropriate in content, construct, criterion, and face validity, and it had high internal consistency and test-retest reliability.

The reliability of the final version of the MAQA that is presented in the paper was checked in Statistica 13.3. The internal consistency of the MAQA was very high (Cronbach's α = .95, N = 1,161, all items were equally reliable) and test-retest reliability at 6–8 weeks was very satisfactory (r = .85, p < .001, N = 71). Internal consistency (α) for individual groups was as follows: non-students .94, STEM students .92, HS students .92, women .94, and men .93. The final version of the MAQA is in the Appendix (Szczygieł, 2020b).

The Single-Item Math Anxiety Scale (SIMA; Núñez-Peña et al., 2014) contains only one question about the general level of math anxiety: "On a scale from 1 to 10, how math-anxious are you?" The level of math anxiety was assessed on a 10-point scale (1 = *no math anxiety*, 10 = *strong math anxiety*). A higher score in the SIMA means a higher level of general math anxiety. The test-retest reliability of the SIMA for a Polish adult sample was r = .72, N = 50 (at 4–6 weeks, in a different sample than the group described in the article).

The Abbreviated Math Anxiety Scale (AMAS; Hopko et al., 2003; Cipora et al., 2015) is a 9-item scale that measures total anxiety, math learning anxiety (e.g., "Watching a teacher work an algebraic equation on the blackboard"), and math testing anxiety (e.g., "Thinking about an upcoming math test one day before") on a 5-point scale (1 = *low anxiety*, 5 = *high anxiety*). The higher the score on the scale, the more intense math anxiety. The reliability of the scale was: for Total α = .90, Learning α = .84, and Testing α = .91.

Math Attitude

Math Attitude Scale for Adults (MASA; Szczygieł, 2020c) is a three-dimensional scale that consists of cognitive (MASA-C; e.g., "Mathematics is important in everyday life"), affective (MASA-A; e.g., "I get nervous when I cannot solve a mathematical problem"), and behavioral (MASA-B; e.g., "I would be happy to take part in a discussion in which one must demonstrate mathematical knowledge") factors. It is a 19-item scale with answers on a 4-point scale (1 = *I strongly disagree*, 2 = *I rather disagree*, 3 = *I rather agree*, 4 = *I strongly agree*). The higher the number of points, the more positive the attitude. Internal consistency for Total, Cognitive (6 items), Affective (6 items), and Behavior (7 items) scales was α = .93, .80, .86, and .90, respectively. Test-retest reliability at 6–8 weeks (N = 71) was: for Total r = .92, Cognitive r = .85, Affective r = .90, and Behavior r = .86.

Math and Polish Self-Concept and Self-Efficacy

Math Self-Concept (MSC) was measured using one question. The participants were asked to subjectively rate their competencies in mathematics using a 10-point scale: "On a scale from 1 to 10, how would you evaluate your current skills in mathematics?" (1 = *definitely low competencies*; 10 = *definitely high competencies*). Test-retest reliability at 4 to 6 weeks was r = .79 (N = 50, a different sample than the group presented in the article).

Math Self-Efficacy (MSE) was assessed by posing a question about the participants' belief in their chance of performing math tasks correctly: "On a scale from 1 to 10, rate your chance of performing tasks in mathematics correctly, e.g., arithmetical and geometrical tasks." The participants expressed their beliefs on a 10-point scale (1 = *I definitely will not be able to complete the task*; 10 = *I definitely will be able to complete the task*). Test-retest reliability at four to six weeks was r = .76 (N = 50, a different sample than the group presented in the article).

Polish Self-Concept (PSC) was assessed on a 10-point scale using one question about the participants' current skills in Polish: "On a scale from 1 to 10, how would you evaluate your current skills in Polish?" 1 = *definitely low competencies*; 10 = *definitely high competencies*. Test-retest reliability for Polish adults at four to six weeks was r = .82 (N = 50, a different sample than the group presented in the article).

Polish Self-Efficacy (PSE) was measured with one question: "On a scale from 1 to 10, rate your chance of performing tasks in Polish correctly, e.g., essay writing, text analysis". The participants answered the question on a 10-point scale (1 = *I definitely will not be able to complete the task*; 10 = *I definitely will be able to complete the task*). Test-retest reliability at four to six weeks was r = .70 (N = 50, a different sample than the group presented in the article).

Math and Polish Achievement

Grades. The respondents were asked to provide their most frequent grades (from 1 to 6) obtained in Mathematics and Polish in elementary school, junior high school, and high school.

High School Exit Exam (Matura) is a national standardized test conducted in Poland on completion of high school. The participants were asked to indicate the results (from 0 to 100%) they obtained in Mathematics and Polish (provided they were examined).

Procedure

The validation study was carried out with paper and pencil (non-students and HS students) or in an online form (STEM students) from 2017 to 2019. Participants were recruited from universities and public schools in face-to-face meetings and via social media. All scales were administered in Polish. The participants completed the questionnaires in the following order: math and Polish achievement, math and Polish self-concept (MSC, PSC), math and Polish self-efficacy (MSE, PSE), math attitude (MASA), math anxiety (AMAS, SIMA, MAQA), trait and state anxiety (STAI), and gender and profession. Not all respondents answered all of the questions. Therefore, there are differences in the amount of data collected from each questionnaire (see Electronic Supplementary Material, ESM 1, Table E-A, Szczygieł, 2020b).

Results[1]

Structure and Descriptive Statistics of the MAQA

In the first step of the analysis, construct validity was established. To check whether the structure of the MAQA was as unidimensional as it was theoretically assumed, confirmatory factor analysis (CFA) was performed in *lavaan* (R package; Rosseel, 2012). Since the assumption of multivariate normality was violated (Mardia's test: skewness 11,589.04, p < .001, kurtosis 160.12, p < .001) and variables were ordinal, the Diagonally Weighted Least Squares estimator (DWLS) was applied (Mîndrilă, 2010). I relied on common fit indices to evaluate the model-to-data fit: chi-square (χ^2), the root mean square error of approximation (RMSEA), the standardized root mean squared residual (SRMR), the comparative fit index (CFI), and the Tucker-Lewis index (TLI). I adopted the following model evaluation criteria: RMSEA and SRMR smaller than .08, CFI and TLI above .95 (Hu & Bentler, 1999; Kline, 2016). Although the tested unidimensional model did not obtain the required chi-square value, $\chi^2(152)$ = 375.23, p < .001, N = 1,161, the model-to-data fit indices were very high: RMSEA = .036 (90% CI [.031, .040]), SRMR = .057, CFI = .99, and TLI = .99. All factor loadings were standardized, significant, and their values varied from .59 to .77. Therefore, factor loadings are above the minimum recommended value > .40 (Costello & Osborne, 2005). Item-scale correlations varied from .63 to .77 (see Table 1).

Subsequently, the R package *lavaan* (Rosseel, 2012) was used to test whether the MAQA had the same measurement characteristics across all groups involved in the study. The decision to confirm each level of equivalence was

[1] I report how I determined my sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. If I use inferential tests, I report exact p values, and effect sizes.

Table 1. Confirmatory factor analysis (CFA) and descriptive statistics in the whole group

Item	*M* (*SD*)	Skewness	Kurtosis	Item-scale Pearson's *r* correlation	β CFA
1. Calculation of the average level of fuel consumption of a car.	1.65 (0.86)	1.19	0.59	.65	.62
2. Calculation of the amount of money that should be deposited every month to save 13,500 PLN within 3 years.	1.36 (0.67)	2.04	4.07	.63	.59
3. Determine whether the given definition of the number π is true: the ratio of the circumference of a circle to its diameter.	1.86 (0.94)	0.79	−0.43	.77	.77
4. Calculation of how many square meters of bathroom tiles should be purchased.	1.49 (0.79)	1.60	1.82	.71	.69
5. The answer to the question of whether a prism can have 37 vertices.	1.90 (0.97)	0.72	−0.59	.77	.76
6. Give an example of a sequence of natural numbers.	1.51 (0.81)	1.55	1.58	.71	.70
7. Calculation of the surface area of a sphere.	1.72 (0.91)	1.08	0.15	.74	.72
8. The answer to the question, in how many ways can 6 people be seated at a table.	1.64 (0.82)	1.16	0.66	.64	.62
9. Finding who is the tallest if Tom is taller than Jane, Tom is shorter than Peter, Peter is taller than Monica, and Monica and Jane are the same height.	1.51 (0.81)	1.55	1.62	.70	.68
10. Presentation of the geometrical interpretation of the equations $y = -x$, $y = x + 2$.	1.80 (1.01)	0.95	−0.40	.74	.73
11. Calculation of 12% less than 215.	1.39 (0.73)	1.93	3.10	.69	.66
12. Calculation of the length of a route based on knowledge of the traveled time and average speed.	1.66 (0.90)	1.21	0.42	.77	.76
13. Giving an example of the practical application of trigonometric functions.	2.22 (1.1)	0.29	−1.28	.75	.73
14. Indication of the number of bisectors in a triangle.	1.97 (1.06)	0.68	−0.88	.76	.75
15. Calculation of average monthly earnings in the case of a 5% increase.	1.44 (0.74)	1.74	2.53	.68	.64
16. Indication of the prime number among the numbers 276, 277, 278, 279, 280.	1.60 (0.85)	1.29	0.71	.71	.70
17. Determining the probability that today is Wednesday.	1.82 (0.97)	0.84	−0.48	.68	.66
18. Dividing the number 10,179 by 13.5 without using a calculator.	2.06 (1.07)	0.53	−1.04	.67	.64
19. Calculation of the sum $2396\frac{7}{9} + 6,725$.	1.75 (1.01)	1.08	−0.15	.75	.74

Note. N = 1,249; all factor loadings and item-scale Pearson's r correlations are on the level $p < .001$.

based on the following criteria. Because chi-square is sensitive to sample size (Cheung & Rensvold, 2002), the principle that the χ^2/df ratio should be smaller than 3 was applied instead of testing the chi-square difference (Schermelleh-Engel et al., 2003). Configural invariance was tested in a model with no equality constraints imposed based on common model fit indices, metric invariance was examined by fitting models where factor loadings on respective items were constrained to be equal across the groups, and scalar invariance was examined based on constraint intercepts to be equal across the groups. The evaluation of metric and scalar invariance models has completed following Chen's (2007) recommendation. The difference of fit indices between nested models in a large sample size (N > 300) should be smaller than .015 for RMSEA, .03 for SRMR, and .01 for CFI and TLI.

First, measurement invariance across gender was checked (see Table 2). The results indicated that the CFA model with the DWLS estimator in women and men obtained the χ^2/df value below 3 and the model fit indices were very satisfactory. Subsequently, a series of nested multigroup CFA models with increasing parameter constraints were used to test configural, metric, and scalar invariance. The results demonstrated a very good data fit for configural invariance across gender. Therefore, in the next step, metric invariance was checked. The results showed that factor loadings on respective items were close to equal across the groups. Although ΔRMSEA equaled .018, ΔSRMR was .015, and ΔCFI and ΔTLI were less than .01. Moreover, it was tested whether the factor loadings met the criteria of the minimum recommended value > .40 in each group (Costello & Osborne, 2005), and the results confirm that there were no items to be removed (all factor loadings ≥ .51, see ESM 1, Table E-B, Szczygieł, 2020b). Given this slight deviation of ΔRMSEA from the criteria, metric invariance may be cautiously rated as

Table 2. Measurement invariance of the MAQA across gender groups

	N	*df*	χ^2	χ^2/df	RMSEA [90% CI]	SRMR	CFI	TLI
Women	820	152	338.54***	2.23	.039 [.033, .044]	.059	.99	.99
Men	322	152	90.14	0.59	0	.073	1.00	1.00
Configural	–	304	428.68***	1.41	.027 [.021, .033]	.060	.99	.99
Metric	–	322	695.64***	2.16	.045 [.041, .050]	.075	.99	.99
Scalar	–	340	795.95***	2.34	.049 [.044, .053]	.077	.98	.98

Note. ***p < .001.

Table 3. Measurement invariance of the MAQA across non-student and student groups

	N	*df*	χ^2	χ^2/df	RMSEA [90% CI]	SRMR	CFI	TLI
Non-students	316	152	234.68***	1.54	.040 [.031, .052]	.080	.99	.99
STEM students	422	152	42.03	0.28	0	.050	1.00	1.00
HS students	423	152	303.30***	2.00	.049 [.041, .057]	.070	.98	.98
Configural	–	456	580.01***	1.27	.027 [.019, .033]	.063	.99	.99
Metric	–	492	917.59***	1.87	.047 [.043, .052]	.090	.98	.98
Scalar	–	528	1,162.33***	2.20	.056 [.051, .060]	.097	.97	.97

Note. ***p < .001.

confirmed. Therefore, in the final step of testing measurement invariance across gender, scalar invariance was checked. The results showed full metric invariance with satisfactory model indices.

Finally, measurement invariance in non-students, STEM students, and HS students was tested (see Table 3). The results of the CFA model with the DWLS estimator demonstrated a very good model-data fit in all groups ($\chi^2/df < 3$, RMSEA and SRMR $\leq$ .08, CFI and TLI > .95). In the first step of measurement invariance testing, a factorial structure was established across the groups. The CFA model had satisfactory properties and the results confirmed configural invariance. In the next step of the analysis, metric invariance was mostly confirmed: the model fit the data well and the comparison of the two models indicated that the change of the goodness-of-fit properties mostly met the criteria and was acceptable: ΔCFI and ΔTLI were less than .01, ΔSRMR = .027, but ΔRMSEA = .020. Then, it was tested whether the factor loadings met the criteria of the minimum recommended value in each group and the results show that all items were higher than .40 (Costello & Osborne, 2005) (all factor loadings $\geq$ .50, see ESM 1, Table E-B, Szczygieł, 2020b). Finally, scalar invariance was tested across non-students, STEM students, and HS students. Because the model fit was acceptable and the criteria were met, it may be concluded that the items had the same item intercept across the groups and were comparable.

Summing up, the MAQA was verified as unidimensional and mostly met the criteria for configural, metric, and full scalar invariance across gender (women and men) and field of study (non-students, STEM students, and HS students). Therefore, in the final step of the analysis, the descriptive statistics of the MAQA were calculated for the whole sample. The average level of math anxiety was weak (M = 1.70, SD = 0.64, N = 1,218), and its distribution was right-skewed (skewness = .83, W = .91, p < .001) and mesokurtic (kurtosis = .01).

The Validity of the MAQA

The convergent and divergent validity of the MAQA was tested through a series of analyses prepared in Statistica 13.3. First, gender differences were tested with the Mann-Whitney's U-test. A non-parametric test was used because the assumptions of group equivalence and homogeneity of variance (Leven's test: $F(1, 196) = 66.37$, $p < .001$) were violated. The results indicated that women (U_1 = 583,783, N = 870) had a higher level of math anxiety than men (U_1 = 134,418, N = 328): $z = -11.67$, $p < .001$, and that gender differences were large (effect size: $\eta^2 = .11$, Cohen's $d = .72$; Lenhard & Lenhard, 2016).

In the next step of analysis, the MAQA score was compared between non-students (M = 1.85, SD = 0.62), STEM students (M = 1.25, SD = 0.36), and HS students (M = 2.02, SD = 0.62). The groups were unequal (N = 347, 424, 447, respectively). The non-normality of the MAQC in the three subgroups was observed (Shapiro-Wilk: $W_1 = .95$, $p < .001$; $W_2 = .68$, $p < .001$; $W_3 = .98$, $p < .001$), and the assumption of homogeneity of variance was violated (Leven's test: $F(2, 1{,}215) = 89.92$, $p < .001$). Because the assumptions for performing one-way analysis of variance test between the groups were violated, the non-parametric Kruskal–Wallis test was performed, $H(2, 1{,}218) = 397.58$, $p < .001$. The results indicated that non-students (mean of range = 700.19) and STEM students (mean of range = 339.90), HS students (mean of range = 794.83) and STEM students, and HS students and non-students differed significantly from one another in the MAQA score ($p < .001$), and these differences were large (effect size: $\eta^2 = .33$, Cohen's $d = 1.39$; Lenhard & Lenhard, 2016). The results are in accordance with the hypotheses.

Subsequently, a series of Pearson's r correlation analyses between all tested variables were performed in the whole group (N = 335–1302) (see ESM 1, Table E-C, Szczygieł, 2020b). The effect size is provided based on Evans' proposal (1996): $r < .20$ very weak, .20–.39 weak, .40–.59 moderate, .60–.79 strong, and > .80 very strong correlation. To test the convergent and divergent validity of the MAQA, the results of MAQA correlations are presented below (the false discovery rate (FDR) correction of the alpha level was applied based on calculations conducted in the R package).

The MAQA score positively and strongly correlated with math anxiety measured with the SIMA ($r = .73$, $p < .001$) and the AMAS ($r = .71$, $p < .001$). The MAQA result negatively and strongly correlated with math attitude total score ($r = -.71$, $p < .001$), moderately/strongly with math attitude dimensions (cognitive $r = -.49$, affective $r = -.68$, behavior $r = -.64$, all $p < .001$), and strongly with math self-concept ($r = -.70$, $p < .001$), and math self-efficacy ($r = -.75$, $p < .001$). The MAQA score negatively and weakly or moderately correlated with math achievement at various stages of education (elementary school $r = -.44$, junior high school $r = -.24$, high school $r = -.50$, high school exit exam Matura $r = -.54$, all $p < .001$). The MAQA positively and very weakly correlated with state ($r = .13$, $p < .01$) and trait anxiety ($r = .19$, $p < .001$), and it positively and weakly correlated with Polish self-efficacy ($r = .20$, $p < .001$) but not with Polish self-concept ($r = .07$, $p = .07$). A positive and very weak relationship or no relationship at all was observed between the MAQA and Polish achievement (elementary school $r = .10$, $p < .01$, junior high school $r = .05$, $p = .14$, high school $r = .08$, $p < .05$, high school exam Matura $r = .14$, $p < .05$). The results are in accordance with the hypotheses and confirm the convergent and divergent validity of the MAQA.

In the next step, the criterion validity comparison of the MAQA, the SIMA, and the AMAS was conducted for HS students (the only adult group in the study to have filled out all of the scales). Correlation tests were conducted in Statistica 13.3 and the FDR correction of the alpha level was applied (based on calculations conducted in the R package). The comparison of correlation coefficients was done based on Lee and Preacher's (2013) calculator. The correlation ($N = 183$) between the AMAS and the SIMA ($r = .76$, $p < .001$) was significantly stronger ($z = 2.14$, $p = .02$) than the correlations between the MAQA and the AMAS and the MAQA and the SIMA (both results $r = .68$, $p < .001$). Further analysis of the criteria was as follows (see Table 4). The correlations between the MAQA and the AMAS were similar in almost all areas: math attitude, math self-concept, math self-efficacy, math achievement (all grades and exit exam), state and trait anxiety, Polish self-concept, Polish achievement (all grades and exit exam). The MAQA's correlations differed significantly from the AMAS in one area: the MAQA did not correlate with Polish self-efficacy while the AMAS score correlated positively and weakly with it. Although a similar pattern of the significance of correlations between the MAQA and the SIMA with criterion was noted, more significant differences in the strength of correlations with criterion were observed between the MAQA and the SIMA than in the MAQA and the AMAS. The MAQA correlated similarly to the SIMA in the behavior dimension of math attitude, math self-concept, math self-efficacy, math achievement (grades in elementary and junior high school), Polish self-concept, and Polish achievement (grades in elementary, junior high school, high school – nonsignificant result in both cases, exit exam). In comparison to the SIMA, the MAQA produced weaker correlations with math attitude (total score, cognitive and affective dimensions), math achievement (grades in high school and exit exam), state and trait anxiety, and Polish self-esteem. The results showing a similar pattern of the significance of the correlation between the MAQA, the AMAS, and the SIMA and criterion mostly confirm the criterion validity of the MAQA; However, some differences in the strength of the relationship between math anxiety scales and criterion variables were noted (especially in the SIMA).

Discussion

Although the level of math anxiety in adults has been surveyed in numerous studies, the ecological validity of some measurements is debatable. It is particularly important to note that some adults do not participate in formal math education, so they do not have math tests and exams, math teachers, math homework, and so forth. In response to the need for an accurate assessment of math anxiety in adults (Núñez-Peña et al., 2014), a new academic-independent scale was designed. The items relate to the level of math anxiety in the face of mathematical problem solving, but not to math teachers, math lessons, or math tests. The structure, validity, and reliability of the MAQA were

Table 4. Criterion validity comparison of math anxiety measures in HS students

		MAQA	SIMA	AMAS	MAQA vs. SIMA	MAQA vs. AMAS
Variable	Measure	Pearson's *r*			*z*-score (two-tail *p*)	
Convergent validity						
Math attitude	MASA Total	−.71***	−.82***	−.71***	3.71 (< .001)***	0 (1)
Cognitive	MASA-C	−.49***	−.61***	−.49***	2.89 (< .01)**	0 (1)
Affective	MASA-A	−.67***	−.81***	−.71***	4.50 (< .001)***	1.14 (.26)
Behavior	MASA-B	−.60***	−.61***	−.53***	0.25 (.80)	−1.06 (.29)
Math self-concept	MSC	−.54***	−.65***	−.51***	1.53 (.13)	−1.90 (.06)
Math self-efficacy	MSE	−.58***	−.65***	−.53***	1.01 (.31)	−1.66 (.10)
Math achievement	Grade elementary school	−.43***	−.43***	−.40***	−0.65 (.52)	0 (1)
	Grade junior high school	−.49***	−.53***	−.49***	0.92 (.36)	0 (1)
	Grade high school	−.41***	−.52***	−.41***	2.46 (.01)*	0 (1)
	High school exit exam (Matura)	−.54***	−.62***	−.48***	1.97 (< .05)*	−1.38 (.17)
Divergent validity						
State anxiety	STAI X-1	.11	.29***	.14	−3.59 (< .001)***	−.59 (.56)
Trait anxiety	STAI X-2	.13	.32***	.20**	−3.82 (< .001)***	−1.38 (.17)
Polish self-concept	PSC	.12	.12	.12	0 (1)	0 (1)
Polish self-efficacy	PSE	.10	.20*	.22**	−1.97 (< .05)*	−2.37 (< .05)*
Polish achievement	Grade elementary school	−.01	−.04	−.05	0.58 (.56)	0.78 (.44)
	Grade junior high school	−.04	.01	.01	−0.97 (.33)	−0.97 (.33)
	Grade high school	−.08	.04	−.02	−2.33 (< .05)*	−1.16 (.24)
	High school exit exam (Matura)	.07	.08	.09	−0.19 (.85)	−0.39 (.70)

Note. N = 183. Correlations between the MAQA, the SIMA, the AMAS, and other variables presented in Table 4 differ from those presented in the Electronic Supplementary Materials, Table C (Szczygieł, 2020b). This is because a different number of participants (observations) were taken into account in the analysis. ***$p < .001$; **$p < .01$; *$p < .05$.

checked and the satisfactory psychometric properties of the scale were established.

The MAQA is a unidimensional scale with high factor loadings (Hair et al., 2006). The measurement invariance of the scale has been confirmed (Lubiewska & Głogowska, 2018). Regardless of gender or regardless of whether STEM students, HS students, or non-students were tested, the MAQA has a comparable structure and model fit properties. Although the average level of math anxiety in the tested group is low, differences in the intensity of anxiety were revealed across the groups.

The results indicate that women have a higher level of math anxiety than men and that the difference is substantial (Lenhard & Lenhard, 2016). The observed gender difference is consistent with other studies (Ashcraft, 2019; Devine et al., 2012). Large differences in the level of math anxiety were observed between non-students, STEM students, and HS students. In accordance with the assumptions, STEM students have the lowest level of math anxiety, HS students have the highest math anxiety level, and the feelings of non-students fall between the two. These results confirm the high validity of the MAQA because adults studying math-related subjects should feel weak math anxiety and it should be lower than anxiety in adults who chose social sciences or humanities. Moreover, non-students may pursue various professions, either related to STEM or HS. The results are consistent with other studies showing that adults can still feel some anxiety in the face of math problems (Ashcraft, 2019; Hart & Ganley, 2019; Hopko et al., 2003; Núñez-Peña et al., 2014).

The convergent and divergent validity of the scale complies with the formulated assumptions and is very satisfactory. First, the MAQA score highly and positively correlates with general math anxiety (SIMA) and total anxiety in the academic context (AMAS). This result is in line with the expectations because previous studies indicate that math anxiety dimensions highly correlate with each other (Cipora et al., 2015). The proper convergent validity is also shown by the correlations between math problem solving and math attitude. As expected, the higher score in the MAQA, the lower score in the MASA. Especially strong relationships were observed between math anxiety and affective and behavioral domains. It is justified since the affective dimension of math attitude includes, among other things, the feeling of anxiety related to math, and adults with high math anxiety do not like math and do not want to engage in math activities. A moderate negative correlation was observed between the cognitive dimension of math attitude and the MAQA score, which means that people who feel high math problem-solving anxiety also believe that mathematics is less useful. As expected, negative and moderate relations between the MAQA score and math self-concept, math self-efficacy, and math achievement (Eden et al., 2013; Zhang et al., 2019) were observed. When the level of math anxiety measured by the MAQA increases, the levels of self-concept, self-efficacy, and grades decrease at all stages of education and the high school exit exam (Matura). Such results indicate that math problem-solving anxiety in adulthood is highly related to experiences rooted in the school context. This is not surprising because the experience gained over many years of schooling shapes the level of math anxiety in adulthood (Hart & Ganley, 2019). It is worth noting that correlation coefficients between the MAQA and math grades differ from each other (weak/moderate). This may be explained by the fact that school grades depend on many other factors than math anxiety (e.g., teachers' requirements). The strong relationship between math self-concept and math problem-solving anxiety produced a comparable correlation coefficient to that of math self-efficacy and math problem-solving anxiety. This was expected because these two are related theoretical constructs (Durrani & Tariq, 2009; Eden et al., 2013). These results confirm the convergent validity of the MAQA.

Moreover, the divergent validity of the MAQA was additionally confirmed. The MAQA score positively and weakly correlated with state and trait anxiety, which means that the MAQA measures a specific kind of anxiety (Ashcraft, 2019; Cipora et al., 2015; Hopko et al., 2003) related to mathematics. The feelings of anxiety as a state and trait only partially explain the level of anxiety related to mathematical problems. Math problem-solving anxiety is not academic-dependent, but it is specific to mathematics. As expected, there is no relationship or weak correlation between the MAQA score and Polish self-concept, self-efficacy, and achievement (grades and the high school exit exam) (Cipora et al., 2015). Polish self-efficacy was positively and weakly related to the MAQA score, which may be explained by the fact that the higher math anxiety, the higher the engagement in school subjects other than math. Such a relationship may persist in adulthood because beliefs about one's own efficacy may be confirmed in everyday activity. It was expected that Polish self-esteem and the MAQA would produce a comparable correlation coefficient to that of Polish self-concept and the MAQA, but Polish self-concept was not related to math problem-solving anxiety. An irregular pattern of correlations was observed also between the MAQA and Polish achievement (non-significant or very weak relationships). Therefore, it is likely that the relationship was moderated or mediated by factors that were not analyzed in the study. Nevertheless, the strength and direction of the relationships were in line with the preliminary assumptions. Summing up, the results mostly indicate that math anxiety in adulthood is specific to mathematics, and they confirm the convergent and divergent validity of the scale.

To ensure that the MAQA is as valid as the AMAS and the SIMA, criterion validity was established. However, it should be noted that analyses were conducted only in HS students because only in this sample all of the tests were administered in a sufficiently large number (therefore, some correlation coefficients in the HS group differ from those in the whole group). Correlations between the MAQA, the SIMA, and the AMAS were compared with other variables and the strength of the correlations between these scales and criterion differ significantly in some points. However, the pattern of significance and effect size of the correlations between criterion variables and the MAQA is almost the same as in the AMAS and the SIMA, which indicates that the scales are mostly comparable. Interestingly, the strength of the correlations between the MAQA and other variables is more similar to the strength of the correlations between the AMAS and these variables than the SIMA and these variables. It is likely that the meaning of one question in the SIMA is interpreted by adults in various ways, which in turn affects the results. Moreover, the MAQA score is similar to the AMAS in that it is related to state and trait anxiety (a very weak effect/no effect) while the SIMA is related more strongly to state and trait anxiety. This suggests that the SIMA score is more rooted in general anxiety than the MAQA and the AMAS scores. The pattern and strength of the correlations between the MAQA, the SIMA, the AMAS, and criterion variables suggest that the MAQA has even a somewhat better divergent validity than the SIMA and the AMAS in some points. The MAQA score in the HS group is not related to Polish self-concept, self-efficacy, and achievement, whereas the SIMA and the AMAS scores are not connected with Polish self-concept and achievement, but are related to Polish self-efficacy. Moreover, it should be noted that the correlation between the AMAS and the SIMA is significantly stronger than the correlations between the MAQA and the AMAS and the MAQA and the SIMA, which suggests (based on item content) that the AMAS and the SIMA are more school-related than the MAQA. On the other hand, the following result weakens the criterion validity of the MAQA: observed in the whole group, weak and positive correlations between the MAQA and state and trait anxiety disappeared when data were analyzed in the HS group, while such a relationship still exists in the case of the SIMA (weak effect) and the AMAS (weak effect in trait anxiety). The differences in the significance of the correlations of the MAQA and state and trait anxiety result from the number of observations because the values of the correlations are very similar. Nevertheless, the specificity of the MAQA including state and trait anxiety needs more studies in the future.

The construct, convergent and divergent, and criterion validity of the MAQA is very satisfactory. The internal consistency and test-retest reliability of the scale is high. However, one might ask what added value the MAQA brings to research. I postulate that its biggest advantage is that it is the most apt in ecological terms. The MAQA was constructed on the premise that the content of the items must be adequate to adults and their situation. Indeed, the results of the face validity study (see Face Validity Study: Scope of Application Math Anxiety Measurements [MAQA, SIMA, AMAS] in Electronic Supplementary Materials, Szczygieł, 2020b) conducted among the adults who assessed the usefulness of various math anxiety scales (MAQA, SIMA, AMAS) indicated that the MAQA was evaluated as better than the SIMA and the AMAS for non-students (math teachers, preschool teachers, early education teachers, parents of young children). Moreover, the MAQA was evaluated as worse than the AMAS and the SIMA for elementary students and worse than the AMAS for high school students. No differences in the evaluation of the MAQA and the SIMA's scope for HS, STEM, and high school students were observed. Nevertheless, the MAQA was evaluated as worse than the AMAS in STEM students but better than the AMAS in HS students. These results confirm the fact that adults intuitively distinguish which items make sense to which age/profession.

Finally, methodological limitations also should be pointed out: the data were collected in different ways in online and paper and pencil forms; more women than men were examined, which was due to the specifics of the tested sample; the validation study was completed mostly in students and validation proofs were collected at the self-report level; the criterion validity of the MAQA was tested only in HS group which limited the strength of the results. Therefore, the MAQA should be tested in the future in more standardized conditions, and the validation proofs should also be collected in non-student samples and in more gender-equivalent groups. The validation study should also include observation and behavioral methods. High correlations between all items and the high reliability of all items suggest that a shortened version of the MAQA may be developed in the future. It may be desirable to shorten the scale, especially if the participants receive a large number of tests to complete (Hopko et al., 2003).

Conclusion

Math anxiety in adults is usually measured with scales fitted for children's and adolescents' learning context (e.g., AMAS, MARS). Although such measurements are adequate for STEM students, using the AMAS or the MARS for non-students or HS students is controversial. An ecologically valid scale intended for various groups of adults was presented. The MAQA scale measures anxiety related to math problem-solving in a valid and reliable way. Therefore, this scale may be recommended for use in future research on non-students, HS students, and STEM students.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/1015-5759/a000646

ESM 1. Tables E-A to E-D: Number of Participants Completing the Questionnaires in the Validation Study; MAQA Factor Loadings by Gender and Group; Correlation Matrix between All Variables. This file contains also details about the Face Validity Study (Scope of Application of Math Anxiety Measurements: MAQA, SIMA, AMAS) and the full version of the Math Anxiety Questionnaire for Adults (MAQA).

References

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

Ashcraft, M. H. (2019). Models of math anxiety. In I. C. Mammarella, S. Caviola, & A. Dokwer (Eds.), *Mathematics anxiety. What is known and what is still to be understood* (pp. 1–19). Routledge.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107*(5), 1860–1863. https://doi.org/10.1073/pnas.0910967107

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

Casad, B. J., Hale, P., & Wachs, F. L. (2015). Parent-child math anxiety and math-gender stereotypes predict adolescents' math education outcomes. *Frontiers in Psychology, 6*, Article 1597. https://doi.org/10.3389/fpsyg.2015.01597

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 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, 13*, 531–542. https://doi.org/10.1207/S15328007SEM0902_5

Cipora, K., Szczygieł, 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*, Article 1833. https://doi.org/10.3389/fpsyg.2015.01833

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*, 1–9.

Devine, A., Fawcett, K., Szucs, 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*(33), 1–9.

Durrani, N., & Tariq, V. (2009). Relationships between undergraduates' mathematics anxiety and their attitudes towards developing numeracy skills and perceptions of numerical competence. In *2nd International Conference of Education, Research and Innovation, November 16–18, 2009, Madrid, Spain*.

Eden, C., Heine, A., & Jacobs, A. M. (2013). Mathematics anxiety and its development in the course of formal schooling – a review. *Psychology, 4*(6A2), 27–35. https://doi.org/10.4236/psych.2013.46A2005

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Brooks/Cole Publishing.

Ganley, C. M., Schoen, R. S., LaVenia, M., & Tazz, A. M. (2019). Construct validation of the Math Anxiety Scale for Teachers. *AERA Open, 5*(1), 1–16. https://doi.org/10.1177/2332858419839702

Hadley, K. M., & Dorward, J. (2011). The relationship among elementary teachers' mathematics anxiety, mathematics instructional practices, and student mathematics achievement. *Journal of Curriculum and Instruction, 5*(2), 27–44. https://doi.org/10.3776/joci.2011.v5n2p27-44

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

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

Hart, S. A., & Ganley, C. M. (2019). The nature of math anxiety in adults: Prevalence and correlates. *Journal of Numerical Cognition, 5*(2), 122–139. https://doi.org/10.5964/jnc.v5i2.195

Hopko, D. R., Mahadeven, 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., & Bentler, P. M. (1999). Cutoff criteria for fi t 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

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

Hunt, T. E., & Sari, M. H. (2019). An English version of the Mathematics Teaching Anxiety Scale. *International Journal of Assessment Tools in Education, 6*(3), 436–443. https://doi.org/10.21449/ijate.615640

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. The Guilford Press.

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology, 9*, 2207–2230. https://doi.org/10.4236/psych.2018.98126

Lee, I. A., & Preacher, K. J. (2013, September). *Calculation for the test of the difference between two dependent correlations with one variable in common* [Computer software]. http://quantpsy.org

Lenhard, W., & Lenhard, A. (2016). *Calculation of effect sizes*. Psychometrica. https://www.psychometrica.de/effect_size.html. https://doi.org/10.13140/RG.2.2.17823.92329

Lubiewska, K., & Głogowska, K. (2018). Zastosowanie równoważności pomiarowej w badaniach psychologicznych [Implementation of measurement equivalence analysis in psychological research]. *Polskie Forum Psychologiczne, 23*(2), 330–356. https://doi.org/10.14656/PFP20180207

Maloney, E. A., Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2015). Integrational effects of parents' math anxiety on children's math achievement and anxiety. *Psychological Science, 26*(9), 1480–1488. https://doi.org/10.1177/0956797615592630

Martin, C. R., & Hollins Martin, C. J. (2017). Minimum sample size requirements for a validation study of the Birth Satisfaction Scale-Revised (BSS-R). *Journal of Nursing and Practice, 1*(1), 25–30. https://doi.org/10.36959/545/358

Mîndrilă, D. (2010). Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal of Digital Society, 1*(1), 60–66.

Newton, P. E., & Shaw, S. D. (2019). *Validity in educational and psychological assessment*. Sage.

Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (2014). The Single-Item Math Anxiety Scale: An alternative way of measuring mathematical anxiety. *The Journal of Psychoeducational Assessment, 32*(4), 306–317. https://doi.org/10.1177/0734282913508528

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

Ramirez, G., Hooper, S. Y., Kersting, N. B., Ferguson, R., & Yeager, D. (2018). Teacher math anxiety relates to adolescent students' math achievement. *AERA Open, 4*(1), 1–13. https://doi.org/10.1177/2332858418756052

Richardson, F. C., & Suinn, R. M. (1972). The Mathematics Anxiety Rating Scale: Psychometric data. *Journal of Counseling Psychology, 19*(6), 551–554.

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

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, 8*(2), 23–74.

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy, 2*(4), 419–436. https://doi.org/10.1207/S15327078IN0204_02

Spielberger, C. D., Strelau, J., Tysarczyk, M., & Wrześniewski, K. (2011). *Podreôcznik STAI – Inwentarz Stanu i Cechy Leôku STAI* [The State-Trait Anxiety Inventory Manual]. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.

Soni, A., & Kumari, S. (2017). The role of parental math anxiety and math attitude in their children's math achievement. *International Journal of Science and Mathematics Education, 15*(2), 331–347. https://doi.org/10.1007/s10763-015-9687-5

Szczygieł, M. (2020a). When does math anxiety in parents and teachers predict math anxiety and math achievement in elementary school children? The role of gender and grade year. *Social Psychology of Education, 23*, 1023–1054. https://doi.org/10.1007/s11218-020-09570-2

Szczygieł, M. (2020b, December 17). *Supplementary Materials to the paper Not only reliability! The importance of the ecological validity of the math anxiety questionnaire for adults*. https://doi.org/10.17605/OSF.IO/E2479

Szczygieł, M. (2020c). *Sukcesy i porażki w uczeniu sieô matematyki. Uwarunkowania osiągnięć matematycznych uczniów edukacji wczesnoszkolnej* [Successes and failures in learning mathematics. Determinants of mathematical achievements in early childhood education]. Libron.

Wu, S. S., Barth, M., Amin, H., Melcarne, V., & Menon, V. (2012). Math anxiety in second and third graders and its relation to

mathematics achievement. *Frontiers in Psychology, 3*(162), 1–11. https://doi.org/10.3389/fpsyg.2012.00162
Vahedi, S., & Farrokhi, F. (2011). A confirmatory factor analysis of the structure of abbreviated Math Anxiety Scale. *Iranian Journal of Psychiatry, 6*, 47–53.
Zhang, J., Zhao, N., & Kong, Q. P. (2019). The relationship between math anxiety and math performance: A meta-analytic investigation. *Frontiers in Psychology, 10*, Article 1613. https://doi.org/10.3389/fpsyg.2019.01613

History
Received January 18, 2020
Revision received January 25, 2021
Accepted February 8, 2021
Published online April 22, 2021
EJPA Section / Category Educational Psychology

Open Data
I confirm that there is sufficient information for an independent researcher to reproduce all of the reported results, including the codebook if relevant. I confirm that there is sufficient information for an independent researcher to reproduce all of the reported methodologies. The data is available (Szczygieł, 2020b). This study was not preregistered with an analysis plan.

Funding
This work was partially supported by the National Science Centre (Poland; grant number 2015/19/N/HS6/00791).

ORCID
Monika Szczygieł
https://orcid.org/0000-0001-6544-0734

Monika Szczygieł
Institute of Psychology
Pedagogical University of Krakow
Podchorazych 2
30-084 Krakow
Poland
monika.szczygiel@up.krakow.pl