ABSTRACT

Validity, Reliability, and Gender Invariance of the Abbreviated Math Anxiety Scale (AMAS) in Middle School Students

Lauren Adelyn Cohen, Psy.D.

Mentor: Christine A. Limbers, Ph.D.

Math anxiety is a common form of state anxiety that is associated with poorer math performance and achievement in children, adolescents, and young adults. It is also associated with avoidance of advanced education and career paths in STEM-related fields and is disproportionately higher in females than males across the lifespan. The Abbreviated Math Anxiety Scale (AMAS) is a brief self-report measure of math anxiety comprised of two subscales that has been shown to be a reliable and valid measure of math anxiety for elementary, high school, and college-aged students. The AMAS also demonstrates factorial invariance across gender in these populations, indicating that it can be used to compare mean score differences in males and females. Despite the importance of the middle school years on the trajectory of math anxiety, the psychometric properties of the AMAS have not yet been examined in a middle school population. The purpose of the current study was to address gaps in the literature by examining the reliability, validity, and gender invariance of the AMAS in middle school students.

A group of 604 students from two middle schools in the Southern United States completed the AMAS in person or online, as well as measures of math anxiety, test anxiety, worry, attitudes towards math, positive affect and career interest in STEM fields. Confirmatory factor analyses and multigroup confirmatory factor analyses were conducted to assess the factor structure of the AMAS and test for factorial invariance across gender. Internal consistency reliability was assessed and correlations between the AMAS and other measures were examined to assess for convergent, and divergent validity.

The AMAS demonstrated good internal consistency reliability, convergent and divergent validity, and factorial invariance across boys and girls in the middle school sample. A bifactor model provided a good fit for the data and an improved fit over unidimensional and bidimensional models. Results from the current study suggest that the AMAS is a valid and reliable measures of math anxiety for middle school students and can be used to measure differences in math anxiety between boys and girls in this population.

Validity, Reliability, and Gender Invariance of the Abbreviated Math Anxiety Scale (AMAS) in Middle School Students

by

Lauren Adelyn Cohen, B.A., M.S.C.P.

A Dissertation

Approved by the Department of Psychology and Neuroscience

____

Charles A. Weaver III, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Doctor of Psychology

Approved by the Dissertation Committee

____

Christine A. Limbers, Ph.D., Chairperson

____

Thomas A. Fergus, Ph.D.

____

Keith P. Sanford, Ph.D.

____

Alisha Wray, Ph.D.

____

Julie K. Ivey, Ph.D.

Accepted by the Graduate School

August 2021

____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Christine Limbers, for providing me with steadfast support and mentorship throughout my six years of graduate training. Dr. Limbers, your guidance, warmth, and commitment to training have helped me develop the clinical and research skills necessary to be a successful pediatric psychologist, and for this I am forever grateful. I would also like to extend gratitude to my other dissertation committee members, Dr. Thomas Fergus, Dr. Keith Sanford, Dr. Alisha Wray, and Dr. Julie Ivey for your guidance and feedback throughout this process.

To my cohort, thank you for always being there to support me and lift me up; I will always treasure our friendship. To mom, dad, and Jacky, thank you for always supporting my dreams and teaching me the importance of hard work, dedication, service, and the pursuit of learning. You have and continue to be the wonderful role models who helped me become the person I am today.

To Mabel and Milton, thank you for always providing me with warmth, comfort, humor, and joy even during the most challenging of times. And finally, to Michael, thank you for being my rock, my sounding board, my cheerleader, and most importantly, my friend. Thank you for always encouraging me to chase my dreams and for joining me on this journey.

CHAPTER ONE

Introduction

Definitions of Math Anxiety

In the 1950's, researchers began investigating the existence of a form of anxiety that was specific to the field of mathematics (Ashcraft & Ridley, 2005). The earliest published research on this topic includes Gough's (1954) study on "mathemaphobia" in female pre-term teachers and Dreger and Aiken's (1957) examination of "numerical anxiety" in undergraduate students. Academic interest in anxiety related to mathematics surged in the 1970's with the publication of Richardson and Suinn's (1972) seminal article on the measurement of math anxiety, "The Mathematics Anxiety Rating Scale: Psychometric Data" (Ashcraft & Ridley, 2005). In what has now become the most commonly cited definition, Richardson and Suinn defined the construct of math anxiety as "anxiety associated with the single area of the manipulation of numbers and the use of mathematical concepts," which, "involves feelings of tension of anxiety that interfere with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations." Richardson and Suinn further suggest that math anxiety may "be a contributor to tensions during routine or everyday activities, such as handling money, balancing bank accounts, evaluating sales prices, or dividing workloads."

After the publication of Richardson and Suinn's article, several other authors attempted to further define the construct known as "math anxiety". Fennema and

Sherman (1976) defined math anxiety as “feelings of anxiety, dread, nervousness, and associated bodily symptoms related to doing mathematics”, while D’Ailly and Bergering (1992) similarly defined it as a “fear and apprehension to specific math-related situations”. Hart and colleagues (2016) also conceptualized math anxiety as an emotional response to mathematics, suggesting that it is a “negative emotional reaction to situations involving math performance or the thought of math performance”. In a more informal and succinct definition, Ashcraft and Ridley (2005) wrote that math anxiety is a “negative reaction to math and to mathematical situations” in both academic and everyday settings.

CHAPTER TWO

Literature Review

The Relationship Among Math Anxiety, General Anxiety, and Test Anxiety

The literature overwhelmingly supports the differentiation of math anxiety, test anxiety, and general anxiety into separate, yet highly related constructs (Dew et al., 1983; Hill et al., 2016; Suárez-Pellicioni et al., 2016). In support of this differentiation of constructs, measures of math anxiety correlate more strongly with one another than with measures of general anxiety or test anxiety (Ashcraft & Ridley, 2005; Dowker et al., 2016; Suárez-Pellicioni et al., 2016), with studies reporting large correlations ranging from .50 to .85 among different measures of math anxiety (Ashcraft & Ridley, 2005; Cohen, 1988; Dew et al., 1983; Hopko et al., 2003). Measures of math anxiety demonstrate smaller correlations ranging from .30 to .69 with measures of test anxiety (Cohen, 1988; Devine et al., 2012; Dowker et al., 2016; Hembree, 1990; Hopko et al., 2003; Kazelskis et al., 2000; McAuliffe & Trueblood, 1986; O'Leary et al., 2017), and .26 to .57 with measures of general anxiety (Cohen, 1988; Hembree, 1990; Hopko et al., 2003; McAuliffe & Trueblood, 1986; O'Leary et al., 2017; Wang et al., 2014), suggesting that math anxiety is a separate construct from test and general anxiety.

In an attempt to investigate the relationship between math anxiety, test anxiety, and general anxiety, McAuliffe and Trueblood (1986) conducted a factor analysis using principal components analysis on measures of the three constructs to investigate overlap of factor structure. The authors found that the three scales loaded onto separate factors, and that scores on the measures of general anxiety and test anxiety only predicted approximately 18% and 16% of variance on the math anxiety measure, respectively. In a meta-analysis, Hembree (1990) found that measures of test anxiety predicted 37% of the variance in measures of math anxiety, whereas alternate measures of math anxiety and test anxiety have demonstrated up to 72% of shared variance (Ashcraft & Ridley, 2005).

Math anxiety is often viewed as a form of 'state' anxiety, in which an individual experiences anxiety during specific situations, rather than 'trait' anxiety, which refers to overall feelings regarding math (Ashcraft & Ridley, 2005; Bieg et al., 2015). Although studies suggest that math anxiety is significantly related to both state and trait anxiety (e.g., McAuliffe & Trueblood, 1986), measures of math anxiety tend to correlate more highly with measures of state anxiety (r = .26 - .52) than with measures of trait anxiety (r = .28 - .51; Hembree, 1990; Hopko et al., 2003; Plake & Parker, 1982).

Results from studies using fMRI data suggest that math anxiety activates many of the same brain regions as fear, general anxiety, and specific anxiety disorders (Artemenko et al., 2015; Suárez-Pellicioni et al., 2016). In their meta-analysis of neuroimaging research of math anxiety, Artemenko and colleagues (2015) found that adults with high levels of math anxiety demonstrate increased activation of the pain network of the insula prior to completing math tasks. Further, relative to children with low levels of math anxiety, children with high levels of math anxiety show a) greater activation of the amygdala during math tasks, suggesting a fear response to math, b) greater connectivity from the amygdala to the ventromedial prefrontal cortex, suggesting activation of compensatory mechanisms, and c) reduced connectivity to the bilateral superior parietal lobule, suggesting a performance deficit (Artemenko et al., 2015).

Findings also suggest that math anxiety is associated with less activation of the dorsolateral prefrontal cortex independent of math performance, with studies demonstrating improvements in math performance following increased emotional processing in this region (Artemenko et al., 2015). This research suggests that math anxiety reduces processing efficiency during math tasks by increasing the amount of mental effort required for emotional regulation (Artemenko et al., 2015).

The Association Between Math Anxiety and Performance

Math anxiety appears to be relatively common, as estimates suggest that 15-20% of the population may experience high levels of math anxiety (Ashcraft & Ridley, 2005). Children report experiencing math anxiety and demonstrate related performance impairments in math as early as elementary school (Maloney & Beilock, 2012; Ramirez et al., 2013, 2016). Studies suggest that levels of math anxiety increase during middle school, peak in early high school, and remain stable through college (Ahmed, 2018; Hembree, 1990; Luo et al., 2009). These findings are concerning given that higher levels of math anxiety are associated with a number of negative outcomes for students ranging from elementary school to college age (e.g., Hembree, 1990; Luttenberger et al., 2018; Núñez-Peña et al., 2013). Specifically, higher levels of math anxiety are significantly associated with poorer math performance, calculation ability, math grades, and math-related achievement (Brush, 1978; Hart et al., 2016; Hembree, 1990; Ma, 1999; Núñez-Peña et al., 2013; Passolunghi et al., 2016; Suárez-Pellicioni et al., 2016). Studies suggest small to moderate negative correlations between math anxiety and performance for elementary and secondary school students (r = -.27 to -.36; Devine et al., 2012; Hembree, 1990; Ma, 1999; Ramirez et al., 2016) and moderate correlations for college students (r = -.31; Hembree, 1990).

Regarding specific areas of math performance, math anxiety demonstrates small to moderate negative correlations with computation ability (r = -.25), knowledge of math concepts (r = -.27), math problem solving [r = (-.27) - (-.32)], spatial ability (r = -.29), and abstract reasoning ability (r = -.40; Hembree, 1990; Wang et al., 2014). However, it is important to note that math anxiety does not appear to be strongly associated with IQ or impairments in reading and writing performance (Hembree, 1990; Hill et al., 2016; Passolunghi et al., 2016; Wang et al., 2014).

Research suggests that math anxiety is also associated with increased avoidance of math-related situations, reduced intent to continue with math courses, reduced enrollment in math courses, enrollment in lower-level math courses, poorer attitudes toward math, and overall increased avoidance of educational tracks and career paths involving mathematics, including jobs in the science, technology, engineering, and math (STEM) sectors (Ahmed, 2018; Ashcraft & Ridley, 2005; Brush, 1978; D'Ailly & Bergering, 1992; Hembree, 1990; Núñez-Peña et al., 2013; Suárez-Pellicioni et al., 2016). Given that a) employment growth in STEM-related jobs is substantially higher than in other fields, and b) salaries for jobs in STEM are higher than the national average (Fayer et al., 2017; National Science Board, 2018), math anxiety puts students at a disadvantage for future career paths (Suárez-Pellicioni et al., 2016).

The Reciprocal Relationship Between Math Anxiety and Performance

The effects of math anxiety and math performance appear to be bidirectional (Ashcraft & Ridley, 2005; Carey et al., 2016; Devine et al., 2012; Dowker et al., 2016;

Maloney et al., 2015). In their review, Carey and colleagues (2016) highlight the deficit theory and the debilitating anxiety model as causal models in the reciprocal relationship between math anxiety and performance.

The deficit theory hypothesizes that poor math performance in early childhood contributes to the development of math anxiety during the school years (Carey et al., 2016). Evidence for the deficit theory derives from two sources. First, longitudinal studies of neurotypically developing children indicate that low math achievement predicts future increases in math anxiety (Carey et al., 2016; Ma & Xu, 2004), suggesting that deficits in math performance lead to increased anxiety surrounding math. Second, children with math learning disabilities report higher levels of math anxiety than their neurotypical peers (Carey et al., 2016; Passolunghi et al., 2016; Rubinsten & Tannock, 2010), suggesting that math anxiety stems from pre-existing difficulties with math.

Conversely, the debilitating anxiety model hypothesizes that intrusive thoughts and worries associated with math anxiety lead to impairments in cognitive processing, which subsequently lead to reduced math performance (Carey et al., 2016). Support for this model comes from studies on stereotype threat, temporary working memory deficits, and attentional biases in individuals with high levels of math anxiety (Ashcraft & Ridley, 2005; Carey et al., 2016; Spencer et al., 1999; J. Steele, 2003).

Ambady and colleagues (2001) used the American societal stereotypes that a) women are inferior to men in mathematic ability, and b) individuals from Asian ethnic backgrounds are superior in mathematic ability, to investigate the effect of stereotype threat on math performance in elementary and middle school girls. The authors found that the Asian-American girls in their sample performed significantly worse on a math test relative to controls when they were explicitly made aware of their gender identity. Conversely, girls who were made aware of their Asian-American ethnic identity performed significantly better on the math test relative to girls in the control and gender stereotype groups (Ambady et al., 2001). In accordance with studies suggesting that anxiety mediates the negative effects of stereotype threat on performance (e.g., Spencer et al., 1999; C. M. Steele, 1997) and the debilitating anxiety model, the results suggest that the condition-associated change in math anxiety mediated the effect of stereotype activation on change in math performance for participants in this study.

In further support of the debilitating anxiety model, a number of studies suggest that math anxiety is associated with temporary deficits in working memory. Ashcraft and Faust (1994) reported that undergraduate students with high and low levels of math anxiety did not demonstrate differences in accuracy or speed when completing simple arithmetic problems. However, when computational problems involving more advanced mathematical concepts were introduced, students in the high math anxiety group demonstrated a significantly higher error rate than those in the low math anxiety group. As solving the more advanced math problems required greater working memory capacity than solving the simple problems, the authors hypothesized that the intrusive thoughts associated with math anxiety occupied a greater proportion of the high math anxiety group's working memory capacities, leading to reduced processing efficiency, and, subsequently, poorer performance (Ashcraft & Faust, 1994).

Ashcraft and Kirk (2001) reported similar findings, demonstrating that math anxiety was negatively associated with computation-based working memory span independent of language-based working memory span. Specifically, participants in the high math anxiety group demonstrated higher error rates on math problems when working memory load was high and lower error rates when working memory load was low (similar to those in the low math anxiety group). These findings further suggest that math anxiety is associated with poorer math performance independent of overall working memory capacity.

Ramirez and colleagues (2016) found a similar pattern of results for first and second grade students. Specifically, higher working memory capacity was associated with higher math achievement and the use of more complex problem-solving strategies dependent on working memory in this sample. Children with higher working memory capacity also showed a stronger negative relationship between math anxiety and math achievement than children with lower math anxiety. This finding suggests that the use of more complex working-memory based problem-solving strategies mediates the negative relationship between math anxiety and achievement for children with higher working memory ability. The authors thus concluded that math anxiety interferes with working memory capacity, most adversely affecting children with larger working memory capacities.

In addition to temporary deficits in working memory processing, the relationship between math anxiety and poorer math achievement appears to be partially mediated by attentional biases. Rubinsten and colleagues (2015) used an implicit test of math anxiety in undergraduates to investigate whether attentional biases associated with math anxiety contribute to deficits in cognitive processing. Participants engaged in a visual probe task, during which stimuli differing in emotional valence were flashed across a screen before being replaced by probes. The rate at which participants responded to the probes was then measured and compared across stimulus conditions. Although the two groups did not differ in accuracy rates, Rubinsten and colleagues found that participants with high math anxiety responded more quickly to probes when they were preceded by math-related primes relative to neutral primes, whereas participants with low math anxiety did not differ in their reaction times between probes. The authors argued that the discrepancy in results provides evidence for selective attention bias for math-related information in the high math anxiety group. They further hypothesized that the attentional bias for math-related stimuli in the high math anxiety group leads to difficulty disengaging from negative thoughts about math in order to engage in problem solving.

Suárez-Pellicioni (2015) also demonstrated an attentional bias for math-related stimuli among undergraduate students with high levels of math anxiety using an emotional Stroop task. The authors found that math anxiety levels were significantly positively associated with response times, ratings, and self-reported difference scores for math-related words in the Stroop task. Specifically, participants with high math anxiety took longer to respond to math-related words than neutral words, while participants with low math anxiety did not demonstrate a difference in reaction times between word types. The authors concluded that individuals with high levels of math anxiety demonstrated an attentional bias toward math content, which participants with high math anxiety interpreted as emotionally threatening. Given that increased attentional bias to threatening stimuli is associated with poorer disengagement from stimuli (e.g., Cisler & Koster, 2010), Rubinsten and colleagues' (2015) and Suárez-Pellicioni's (2015) findings suggest that individuals with high levels of math anxiety likely have increased difficulty disengaging from intrusive negative math-related thoughts, which negatively impacts their ability to engage in problem-solving skills, and subsequently leads to poorer math performance.

Math Anxiety, Race, and Ethnicity

To date, few studies have investigated the relationship between math anxiety and race and ethnicity. Obsborne (2001) investigated whether anxiety mediated racial differences in academic achievement. He found that anxiety partially mediated the relationship between race and achievement between White and African American participants and White and Latino participants in a sample of high school students, explaining 38.8% and 41.4% of the effect of race, respectively. In a meta-analysis of the relationship between math anxiety and performance, Hembree (1990) did not find any significant differences in math anxiety between racial groups in a sample of college students. Hart and Ganley (2019) investigated differences in math anxiety between several racial and ethnic groups in a sample of adults. They did not find any significant differences in math anxiety among Asian, Black or African American, White, American Indian or Alaskan Native, Native Hawaiian or Pacific Islander, or multiracial participants. The authors also did not find any differences in Hispanic/Latino and non-Hispanic/Latino participants.

Casanova and colleagues (2021) conducted a longitudinal study of the relationship between math anxiety and math attitudes in first grade and math achievement in fourth grade among a sample of Black and Latinx elementary school students. The authors found that math anxiety in first grade significantly negatively predicted math achievement in fourth grade for girls only. The authors suggested that girls from minority racial and ethnic populations are particularly susceptible to the negative relationship between math anxiety and math performance. Within a sample of Latinx college students, Fernández and colleagues (2021) found that female Latinx students reported more math anxiety than male Latinx students. Among students, having strong math skills predicted lower math anxiety, while strong commitment and class participation in math courses predicted high levels of math anxiety. The authors concluded that interventions targeting improvement in study skills for Latinx participants may be helpful in reducing math anxiety.

Math Anxiety and Psychopathology

Few studies have examined the relationship between math anxiety and clinically significant levels of psychopathology in children and adults. Canu and colleagues (2017) found that college students with ADHD reported significantly more math anxiety than their peers without ADHD, representing a medium effect, (d = .63). Students with ADHD also experienced a significant increase in negative affect following a math test, representing a small effect (d = .38). Conversely, Georgiou and colleagues (2018) found that adolescents with High-Functioning Autism Spectrum Disorder reported lower levels of math anxiety than their neurotypical peers.

Several studies have suggested a positive relationship between math anxiety and learning disorders in math (Carey et al., 2016; Passolunghi et al., 2016; Rubinsten & Tannock, 2010; Wu et al., 2014). For example, Wu and colleagues (2014) found that children with a learning disability in math reported significantly more math anxiety and performed significantly worse on a measure of math achievement than typically developing children and children who were low achieving in math without a learning disorder.

To our knowledge, only one study to date has examined the relationship between math anxiety and internalizing and externalizing symptoms in children (Wu et al., 2014). In a sample of children ages seven- to nine-years-old, Wu and colleagues (2014) found significant positive correlations between scores on a measure of math anxiety and various subscales of the parent-reported Child Behavior Checklist (CBCL), a survey of clinically significant emotional and behavioral problems in children. Specifically, the researchers found positive, statistically significant correlations between math anxiety and the Anxiety/Depression subscale (r = .12), Withdraw/Depression subscale (r = .23), Social Problems subscale (r = .23), Attention Problems subscale (r = .19), Aggression subscale (r = .16), and Other Problems subscale (r = .13) of the CBCL. There were no significant associations between math anxiety and the Somatic Problems, Thought Problems, or Rule Breaking subscales.

Although research examining the relationships between math anxiety and various mental disorders is limited, the studies highlighted above suggest that there appears to be a relationship between math anxiety and some forms of psychopathology. Additional research is needed to clarify the relationships between math anxiety and mental disorders.

Math Anxiety and Gender Differences

Girls tend to report significantly higher levels of math anxiety than boys across school years (Alexander & Martray, 1989; Bieg et al., 2015; Devine et al., 2012; Dowker et al., 2016; Goetz et al., 2013; Hill et al., 2016; Meece et al., 1990; Wang et al., 2014; Wigfield & Meece, 1988). However, gender differences in math performance appear to be very small to negligible (Ashcraft & Ridley, 2005; Devine et al., 2012; Hill et al., 2016; Ma, 1999), with meta-analyses reporting trivial effect sizes of d = 0.05 - 0.07

(Lindberg et al., 2010; Reilly et al., 2015). As the ratio of women to men in STEM careers remains low (Blackburn, 2017), these results suggest that math anxiety, rather than ability, may mediate the relationship between gender and the pursuit of higher education in STEM-related fields.

Measures of Math Anxiety and their Psychometric Properties

The Mathematics Anxiety Scale (MARS)

The Mathematics Anxiety Rating Scale (MARS; Richardson & Suinn, 1972) has been the most commonly used scale for measuring math anxiety since its publication (Alexander & Martray, 1989; Ashcraft & Ridley, 2005; Capraro et al., 2001; Wigfield & Meece, 1988). Several other scales for measuring math anxiety have been developed over the years, but psychometric data regarding these measures is scant. Other than modified versions of the MARS, the most commonly cited early measures of math anxiety include Dreger and Aiken's (1957) Number Anxiety Scale, Fennema and Sherman's (1976) Mathematics Attitude Scales, and Wigfield and Meece's (1988) Math Anxiety Questionnaire (Capraro et al., 2001).

The psychometrics of Richardson and Suinn's (1972) Math Anxiety Rating Scale (MARS) have been extensively studied in several different populations. The MARS is a 98-item scale composed of descriptions of situations involving mathematics, such as "adding two three-digit numbers while someone looks over your shoulder". Respondents rate their level of anxiety elicited by each situation on Likert-type scale ranging from 1 to 5, with 1 indicating "not at all" anxious and 5 indicating "very much" anxious. A total math anxiety score ranging from 98 to 490 is calculated by summing the individual item scores, with higher scores corresponding to higher levels of math anxiety. Subsequent studies have suggested that the MARS can be split into two subscales, which have been labeled the Math Test Anxiety subscale and the Numerical Test Anxiety subscale (Alexander & Cobb, 1987; Brush, 1978; Rounds & Hendel, 1980).

Internal consistency. The MARS demonstrates excellent internal consistency in undergraduate samples, with studies reporting coefficient alphas ranging from .92 to .97 for the total scale (Capraro et al., 2001; Dew et al., 1983; Richardson & Suinn, 1972). Rounds and Hendel reported coefficient alphas of .93 for the Math Test Anxiety subscale and .87 for the Numerical Test Anxiety subscale of the MARS. Overall, the findings suggest that the MARS demonstrates excellent internal consistency.

Test-rest reliability. Richardson and Suinn (1972) reported a seven-week test-retest Pearson product moment correlation coefficient of .85 for the MARS in an undergraduate sample. While Dew, Galassi, and Galassi (1983) found that the measure demonstrated good two-week test-retest reliability in a sample of undergraduate students overall (r = .87), it demonstrated a significantly higher reliability coefficient for men (r = .95) than women (r = .86), suggesting greater stability of scores for men than women. Capraro, Capraro, and Hensen (2001) reported an average test-retest reliability coefficient of .84 across 28 studies. Overall, these findings suggest that the MARS demonstrates good test-retest reliability in adult samples.

Construct validity. As an assessment of construct validity, Richardson and Suinn (1972) collected participant responses to the MARS prior to and following three math anxiety reduction groups at two universities. Mean score reductions were significant for all three groups compared to a control group that did not demonstrate significant change from pre- to post-treatment, suggesting that the MARS demonstrates good construct validity.

The MARS has demonstrated good convergent validity in relation to its correlations with interest in math (r = -.35), math attitudes (r = -.65), highest level of math achieved (r = .44), dislike of mathematics (r = .39), and a test of math ability (r = -.64; Brush, 1978; Resnick et al., 1982; Richardson & Suinn, 1972; Rounds & Hendel, 1980). However, Resnick et al. (1982) found that the MARS had no incremental predictive value for math performance after controlling for course grades, SAT scores, and high school rankings.

The MARS demonstrated stronger correlations with a (reverse-scored) measure of math anxiety (r = -.68) than with a measure of test anxiety (r= .57), suggesting good discriminant validity (Dew et al., 1983). However, the MARS has demonstrated large correlations with the Suinn Test Anxiety Behaviors Scale (r = .65 -75; Suinn, 1969), a measure of test anxiety, calling into question whether the scales measure two separate constructs (Brush, 1978; Rounds & Hendel, 1980). Rounds and Hendel (1980) also implied that the MARS may not measure a construct separate from test anxiety, after finding that the subscales of the MARS did not demonstrate sufficient discriminant validity from the Test Anxiety Inventory (Spielberger et al., 1980).

Factor validity. In an early validation study of a 94-item version of the MARS, Brush (1978) determined that the MARS items loaded onto two factors using a principal components analysis. Forty-five of the items loaded onto the first factor, which he labeled “Problem-Solving Anxiety” and 31 items loaded onto the second factor, which he labeled

“Evaluation Anxiety”. The remaining 18 items did not load onto a factor at a salient magnitude.

Rounds and Hendel (1980) found that the MARS loaded onto two factors using principal components analysis. Items were determined to load onto a factor if they demonstrated loadings of .30 or above. Factor 1, which was composed of 42 items, was labeled “Math Test Anxiety”, as several of the items with salient loadings on this factor were related to anxiety before, during, and after taking math tests. Factor 2, which was composed of 44 items, was labeled “Numerical Anxiety”, as several of the items on the factor were associated with conducting numerical calculations in everyday situations. Three MARS items had factor loadings of .30 on both factors and five items did not have factor loadings of .30 or greater on either factor. The two scales were only moderately correlated at .34, sharing 12% of common variance. There was a significant difference between participant scores on the Math Test Anxiety scale and Numerical Anxiety scale (Cohen’s $d = 2.24$), representing a large effect.

Alexander and Cobb (1987) found a two-factor solution for the MARS using principal components analysis. Factor 1, labeled “Math Test/Course Anxiety”, contained 17 items and accounted for approximately 33% of the variance in item responses. Factor 2, labeled “Numerical Test Anxiety”, contained 7 items accounted for 7% of the variance in scores. The remaining 74 items did not demonstrate salient loadings on either factor. The two factors were only moderately correlated at $r = .32$, sharing 10% of common variance.

Using principal components analysis, Resnick, Viehe, and Segal (1982) determined that a three-factor solution best fit the MARS. Factor 1 was composed of 19 items and accounted for 36% of the variance in scores. It was labeled “Evaluation Anxiety”, as items loading on the factor were associated with evaluation of math ability. Factor 2 was composed of 4 items and accounted for 5% of the variance in scores. It was labeled “Social Responsibility Anxiety”, as items loading on the factor were associated with the real-world application of math in organizations and clubs. Factor 3 was composed of 7 items, which accounted for 4% of the variance in scores. It was labeled “Arithmetic Computation Anxiety”, as the items loading on the factor were associated with everyday situations involving mathematical computation. The remaining 68 items did not demonstrate salient loadings on either factor.

Revised Versions of the MARS

Given the long length of the MARS and inconsistent findings regarding its factor structure, Alexander and Martray (1989) created an abbreviated version of the measure known as the Revised Mathematics Anxiety Rating Scale (RMARS). The authors reduced the 98-item MARS to 69 items and found that a three-factor model fit best. Factor 1, “Math Test Anxiety”, contained 15 items and accounted for 24% of the variance in scores. Factor 2, “Numerical Task Anxiety”, was composed of 5 items accounted for 4% of variance. Factor 3, “Math Course Anxiety”, was composed of 5 items and accounted for 3% of variance. The 44 items that did not load strongly onto the three factors were dropped from the measure.

Baloğlu & Zelhart (2007) conducted a confirmatory factor analysis of Alexander and Martray’s (1989) RMARS using the proposed three factor solution to further investigate the factor validity of the measure. The authors found that the proposed three-factor model was not supported by goodness-of-fit indices under Maximum Likelihood

(ML; CFI = .88, RMSEA = .10) or robust estimation methods (CFI = .88, RMSEA = .09), given the index cutoff values proposed by Hu and Bentler (1999). The authors conducted an exploratory factor analysis (EFA) and found that a different three-factor model best fit the data. The three factors, which retained the same names, accounted for 66.08% of the variance in item scores. Five items with factor loadings smaller than .60 were dropped from the model. The authors conducted a second CFA using the revised model. Although fit for the revised model was improved over the original model, the goodness-of-fit indices under ML (CFI = .92, RMSEA = .09) and robust estimation methods (CFI = .92, RMSEA = .09) did not suggest a good fit for the data. The measure resulted in a similar fit in a second sample under the Maximum Likelihood estimation method (CFI = .91, RMSEA = .09) and an improved fit under the robust estimation method (CFI = .94, RMSEA = .07), although neither met cutoff values for a good fitting model. Baloğlu & Zelhart tested the modified RMARS for factorial invariance across male and female undergraduates. They reported that the model produced a slightly different fit for women (ML: CFI = .91, RMSEA = .06; robust: CFI = .92, RMSEA = .07) than for men (ML: CFI = .92, RMSEA = .09; robust: CFI = .94, RMSEA = .06), but that a comparison of the fit indices demonstrated factorial invariance of the measure across gender.

Like Alexander and Martray (1989), Plake and Parker (1982) created an alternate abbreviated version of the MARS composed of 24 items, also known as the Revised Mathematics Anxiety Rating Scale (MARS-R). The authors chose the majority of the items for the scale from the Math Test Anxiety subscale of the MARS. Using principal components analysis, the authors determined that a two-factor solution best fit the data and accounted for 60% of the variance in scores. Factor 1, “Learning Math Anxiety”, contained 16 items and factor 2, “Mathematics Evaluation Anxiety”, consisted of the remaining eight items. Fit indices were not reported for the model.

Hopko (2003) conducted a confirmatory factor analysis on the MARS-R using the proposed two-factor solution. Goodness-of-fit index results suggested that the two-factor model did not produce a good fit for the data (RMSEA = .09, GFI = .81, AGFI = .78), given suggested fit index cutoffs for assessing adequate model fit (i.e., Hooper et al., 2008; Hu & Bentler, 1999; Shevlin & Miles, 1998). Results of a single factor model produced an even poorer fit than the two-factor model (RMSEA = .10, GFI = .69, AGFI = 63). As a result, Hopko removed items from the measure using a model modification procedure described by Hatcher (1994), until an adequate model fit was achieved. After 12 items were dropped from the model, a two-factor solution composed of the remaining twelve items demonstrated an adequate fit for the data (RMSEA = .05, GFI = .95, AGFI = .93, BCFI = .97). The two factors were highly correlated (r = .72), and the revised model correlated strongly with the original MARS-R (r = .97). Hopko conducted a confirmatory factor analysis in a replication sample to test the modified 12-item measure. Goodness-of-fit indices suggested that the two-factor solution approached an adequate model fit for the data but did not meet cutoff criteria for a good fit for the scale (RMSEA = .06, GFI = .94, AGFI = .92). Hopko tested the revised model for factorial invariance across gender. Fit indices for men (RMSEA = .07, GFI = .94, AGFI = .91, BCFI = .94) and women (RMSEA = .06, GFI = .95, AGFI = .92, BCFI = .96) were similar, suggesting that the model was invariant across gender.

Limitations of the Original and Revised Versions of the MARS

Despite the wide usage of the MARS and its derivations for the measurement of math anxiety, the scale demonstrates a number of substantial limitations. First, the length of the original MARS calls the feasibility of its use as a research instrument into question. At 98 items, the length of the measure creates a large response burden for participants (Rolstad et al., 2011). As research suggests that participant response rates are negatively associated with questionnaire length (Galesic & Bosnjak, 2009; Rolstad et al., 2011; Sahlqvist et al., 2011), the scale's length makes it less useful for data collection than a briefer questionnaire.

Second, the original and abbreviated versions of the MARS lack strong factor validity. As demonstrated above, the factor structures of the MARS, MARS-R, and RMARS differ drastically across studies, even among similar populations. Further, goodness-of-fit indices associated with the "best-fitting" factor solutions for the measures do not meet cutoff criteria for good model fit. The measures have required substantial change in content before the model fit indices have approached acceptable criteria, and only Hopko's (2003) drastically modified version of the MARS-R has demonstrated good model fit.

The Abbreviated Math Anxiety Scale (AMAS)

Development. After finding that his heavily modified version of Plake and Parker's (1982) MARS-R demonstrated a good fitting model, Hopko and colleagues (2003) sought to develop a briefer and more parsimonious version of the measure to assess math anxiety. Using an undergraduate sample, the authors conducted an EFA on item responses to the MARS-R. The authors found that a two-factor solution accounted for approximately 52% of the variance in items scores. They then dropped the 14 items from the scale that did not demonstrate loadings of .70 or above on a single factor. The authors determined that two items were redundant and combined them into a single item, for a total of nine items on the scale. Hopko and colleagues named the scale the Abbreviated Math Anxiety Scale (AMAS).

Hopko and colleagues (2003) conducted an exploratory factor analysis of the nine-item AMAS, finding that a two-factor solution accounted for 70% of variance in item scores. The authors named the two factors “Learning Math Anxiety” and “Math Evaluation Anxiety” after the scales on the MARS-R. The Learning Math Anxiety (LMA) factor was composed of five items, with factor coefficients ranging from .52 - .86. The Math Evaluation Anxiety (MEA) factor was composed of four items, with factor coefficients ranging from .66 - .89. The mean score for the undergraduate sample was 21.1 (SD = 7.0), and female students (M = 21.9, SD = 6.9) reported significantly more math anxiety than male students (M = 19.5, SD = 6.9). The MEA and LMA subscale scores correlated highly with one another (r = .62), although they remained below the .85 cutoff for problematic discriminant validity. The AMAS total score also correlated highly with the LMA (r = .88) and MEA subscales (r = .92).

Hopko and colleagues (2003) conducted a confirmatory factor analysis (CFA) of the AMAS in a replication sample of undergraduate students. Standardized path coefficients ranged from .43 to .86 on the LMA and MEA factors and the two-factor solution produced the following goodness-of-fit statistics: $\chi^2 = 50.81$ ($df = 26$, $p < .001$), RMSEA = .06, GFI = .95, AGFI = .92, BCFI = .96. Based on Hu and Bentler’s (1999) goodness-of-fit cutoff values, the results of the confirmatory factor analysis suggested that the two-factor solution provided a good fit for the data. The two-factor model was also tested for invariance across gender. Results suggested that the two-factor solution produced a good fit for both female [χ^2 = 37.56 (df = 26), p < .001; RMSEA= .05, GFI = .95, AGFI = .91, BCFI = .97] and male undergraduate students [χ^2 = 36.98 (df = 26), p < .001; RMSEA = .07, GFI = .93, AGFI = .90, BCFI = .93].

Factor structure of the AMAS in adult populations. Primi and colleagues (2014) conducted a CFA of an Italian version of the AMAS in high school and undergraduate students. The two-factor model produced a good fit for both the high school sample [χ^2 = 56.46 (df = 26), p < .001; CFI= .96; TLI = .95; RMSEA = .07] and the undergraduate sample [χ^2 =75.10, (df = 26), p < .001; CFI = .95; TLI = .93; RMSEA = .08]. There was a large correlation between the LMA and MEA factors for both the high school (r = .54) and college (r = .55) samples, and standardized path coefficients ranged from .60 to .94 and .51 to .90, respectively. The two-factor model resulted in the following goodness-of-fit index values for males: χ^2 = 74.791 (df = 26), p < .001; CFI = .94; TLI = .92; RMSEA = .08, and the following index values for females: χ^2 = 59.7 (df = 26), p < .001; CFI = .96; TLI = .94; RMSEA = .07. The authors concluded that the comparative fit statistics of the invariance models suggested equality of the scale across genders for both the high school and college samples.

Vahedi and Farrokhi (2011) conducted a CFA of a Persian version of the AMAS in a sample of undergraduate students. The two-factor model produced the following goodness-of-fit statistics: χ^2 = 60.79 (df = 25), p < .001; CFI = .96; TLI = .94; RMSEA = .07, suggesting a good fit for the data. The two factors were highly correlated (r = .67).

The authors conducted a multi-group CFA to evaluate factor invariance across gender. The two-factor model resulted in the following goodness-of-fit index values for females: $\chi^2 = 41.60$ ($df = 25$), $p < 0.02$; CFI = .96; IFI = .95; RMSEA = .06, and the following index values for males: $\chi^2 = 48.88$ ($df = 25$), $p < 0.003$; CFI = .94; IFI= .94; RMSEA = .08. Results of additional analyses suggested that the scale was invariant across gender.

Cipora and colleagues (2015) measured the factor structure a Polish version of the AMAS in undergraduate students. Scores for the LMA and MEA subscales were moderately correlated ($r = .49$) and both subscales were highly correlated with the total AMAS score ($r = .85$, .88, respectively). A confirmatory factor analysis using a two-factor solution produced the following goodness-of-fit statistics: RMSEA = .09, AGFI = .86, suggesting that the model did not provide an adequate fit for the data. The authors conducted a second CFA adding a path from an item assessing math homework-related anxiety on the MEA factor to the LMA factor, with the explanation that homework is related to both learning and evaluation. With the addition of this path, the model produced the following goodness-of-fit statistics: RMSEA = .07, AGFI = .91, suggesting an improved, but still inadequate fit for the data. The results of the CFA suggest that the measure does not provide an adequate fit for a Polish population, comparable to those seen for American, Italian, and Iranian populations (i.e., Hopko et al., 2003; Primi et al., 2014; Vahedi & Farrokhi, 2011).

Factor structure of the AMAS in child populations. Caviola and colleagues (2017) investigated the factor structure of the Italian version of the AMAS in primary school children ages 8 to 11 years old. The AMAS total score demonstrated large correlations with the LMA subscale ($r = .83$) and the MEA subscale ($r = .88$). A CFA using a two-factor solution produced the following goodness-of-fit statistics: $\chi^2 = 153.26$ ($df = 26$), $p < 0.001$; CFI = .93; TLI = .90, RMSR = .07; RMSEA = .07, suggesting a fairly good fit for the data. The authors conducted a multiple-group CFA to investigate the invariance of the model across gender. The model produced a slightly better fit for girls ($\chi^2 = 72.69$ ($df = 26$), $p < 0.001$; CFI = .94; TLI = .92; RMSR= .07; RMSEA = .06) than for boys ($\chi^2 = 104.22$ ($df = 26$), $p < 0.001$; CFI = .91; TLI = .90; RMSR = .08; RMSEA = .08), although additional analyses suggested that measurement error was equivalent across genders.

Carey and colleagues (2017) modified items of the AMAS to make them more applicable to children in the British school system. They conducted a CFA of the modified AMAS (mAMAS) using a two-factor solution to evaluate the model fit for English students ages 8 to 9 and 11 to 13 years-old. All standardized path coefficients for the model were higher than .60, and all parameter estimates were significantly different from zero. The two-factor model produced the following goodness-of-fit statistics: $\chi^2 = 466.95$ ($df = 84$), $p < 0.001$; RMSEA = .07; CFI = .97, suggesting a good fit for the data.

Psychometric Properties of the AMAS. Studies report coefficient alphas for the AMAS in undergraduate populations ranging from .82 to .90 for the total scale, .74 to .85 for the LMA subscale, and .79 to .88 for the MEA subscale (Cipora et al., 2015; Hopko et al., 2003; Primi et al., 2014; Vahedi & Farrokhi, 2011), suggesting that the AMAS demonstrates good internal consistency reliability in adult populations.

Caviola and colleagues (2017) reported alpha coefficients of .77 for the AMAS total scale, .64 for the LMA subscale, .74 for the MEA subscale in primary school students ages 8 to 11 years-old. On the modified AMAS, Carey and colleagues (2017) reported coefficient alphas of .85 for the total scale, .77 for the LMA subscale, and .79 for the MEA subscale for 8- to 9-year-olds, and .86 for the total scale, .80 for the LMA subscale, and .81 for the MEA subscale for 11 to 13 year-olds. Primi and colleagues (2014) reported alpha coefficients of .86 for the AMAS total scale, .81 for the LMA subscale, and .80 for the MEA subscale in high school students ages 14 to 19 years-old. These results suggest that the internal consistency reliability of the AMAS increases with participant age and that the measure demonstrates good internal consistency for children ages 11 years and above.

Hopko and colleagues (2003) reported two-week Pearson product moment correlation coefficients of .85 for the total scale, .85 for the LMA subscale, and .83 for the MEA subscale in an undergraduate population, suggesting good test-retest reliability. Cipora and colleagues (2015) reported four-month test-test reliability coefficients of .71 for the AMAS total scale, .71 for the MEA subscale, and .59 for the LMA subscale, suggesting that the LMA subscale did not demonstrate adequate test-retest reliability in an undergraduate population.

The AMAS has demonstrated strong convergent validity with the MARS-R for the total score (r = .85), LMA subscale (r = .70), and the MEA subscale (r = .81) in undergraduate students (Hopko et al., 2003). It has demonstrated large correlations with measures of test anxiety (r = .52 - .57) and small correlations with measures of general anxiety (r = .22 - .33), attitudes towards math (r = -.48; Hopko et al., 2003; Primi et al., 2014). The AMAS has also demonstrated moderate correlations with high school math grades [r = (-.34) - (-.52)] and number of high school math courses taken (r = -.31;

Hopko et al., 2003), suggesting that the scale demonstrates good convergent validity for adult populations.

Total scores on the AMAS have demonstrated large correlations with measures of test anxiety ($r = .54$) and attitudes towards math ($r = -.53$), as well as moderate correlations with measures of general anxiety ($r = .40$) for high school and primary school students (Caviola et al., 2017; Primi et al., 2014). Hill and colleagues (2016) found that the AMAS demonstrated moderate positive correlations with a measure of general anxiety for students in grades 3 to 5 and moderate negative correlations with math performance for students in grades 6 to 8, although the authors did not report correlation coefficients. Overall, findings suggest that the AMAS demonstrates good convergent validity in children and adolescents.

Carey and colleagues (2017) performed an exploratory factor analysis using the items from the modified AMAS (mAMAS), Children's Test Anxiety Scale (Wren & Benson, 2004), and Revised Children's Manifest Anxiety Scale (Reynolds & Richmond, 1985) to evaluate the discriminant validity of the measures. Results suggested a five-factor solution, with all items from the mAMAS and a single item of the Children's Test Anxiety Scale loading onto a single factor labeled "Math Anxiety". A confirmatory factor analysis of the five-factor model produced the following goodness-of-fit statistics: SRMR = .06, RMSEA = .04, CFI = .94, suggesting an adequate fit for the data. The findings suggest that the mAMAS measures separate constructs from the Children's Test Anxiety Scale and Revised Children's Manifest Anxiety Scale, indicating good divergent validity from test and general anxiety scales in children.

Limitations of Existing Research on the AMAS

The Abbreviated Math Anxiety Scale is the most promising scale currently available for the measurement of math anxiety, given its high internal consistency and test-retest reliability, short-length, strong convergent and discriminant validity, and parsimonious factor structure demonstrating good model fit (Ashcraft & Ridley, 2005; Hopko et al., 2003). Although the psychometric properties of the AMAS have been investigated in American adults, the measure has not yet been validated for use with middle school children in the United States.

Research suggests that math anxiety may increase significantly for almost one-quarter of students over the course of middle school (Ahmed, 2018). This finding is especially alarming, given that students who experience high levels of math anxiety in middle school are less likely than their peers to be employed in STEM jobs as adults (Ahmed, 2018). As STEM-related jobs offer higher salaries and better prospected growth than careers in other industries, math anxiety therefore places students at a disadvantage by the time they finish middle school (Fayer, Lacey, & Watson, 2017; National Science Board, 2018). A significant amount of research and funding has been focused on increasing youth interest in STEM fields in the United States and closing the gender gap in STEM-related careers (Buffington et al., 2016; National Center for Science and Engineering Statistics, 2021; U.S. Department of Education, 2021). In order to identify at-risk individuals, develop efficacious interventions, and monitor treatment outcomes for this population, it is necessary to have a psychometrically sound scale for measuring math anxiety in these populations. Although Carey and colleagues (2017) found a modified version of the AMAS (mAMAS) to be a reliable and valid scale for measuring math anxiety in English children ages 11 to 13 years, there are two major limitations to the study.

Carey and colleagues (2017) did not test for factorial invariance of the measure across girls and boys. Although females report more math anxiety than males on the AMAS, as demonstrated in undergraduate (Cohen's $d = 0.35 - 0.61$), high school (Cohen's $d = 0.49$), primary school (grades 3 to 5; Hedges' $g = .26$), and secondary school (grades 6 to 8) samples (Hedges' g = .28; Cipora et al., 2015; Hill et al., 2016; Hopko et al., 2003; Primi et al., 2014), it is impossible to unambiguously compare these reported rates of math anxiety between boys and girls without first determining whether the measure assesses the construct of math anxiety similarly in both groups (Cheung & Rensvold, 2002; Meredith & Teresi, 2006). Therefore, before the AMAS can be used to measure differences in boys and girls, it must first be shown to measure the same constructs in these two populations. It is important to establish the factorial invariance of a scale across gender before using it in clinical and research contexts, because if measurement invariance is responsible for the score discrepancy between boys and girls rather than true response differences, true need for resources and policy changes may be under or over represented (Meredith & Teresi, 2006). Although factorial invariance across gender has been established for the AMAS among elementary school aged children and adults (Caviola et al., 2017; Hopko et al., 2003; Primi et al., 2014; Vahedi & Farrokhi, 2011), it has not yet been evaluated among middle school age children.

Rodriguez (2016a) suggests that when conducting psychometric analysis of a scale, it is important to determine a) whether specific factors (i.e., subscales) provide predictive validity for external variables above and beyond that of a general factor (i.e., full scale), and b) the strength of the relationship between specific factors and item responses independent of the common factor. Although Carey and colleagues (2017) reported that a two-factor solution produced a good fit for the measure, this correlated-factors solution does not provide information regarding whether the Learning Math Anxiety (LMA) and Math Evaluation Anxiety (MEA) subscales account for additional variance in item responses above and beyond the full scale (F. F. Chen et al., 2006). Therefore, it is unclear whether reporting the subscale scores of the AMAS and using the subscales in statistical analyses provide additional benefit in predicting relationships with other variables beyond the full scale score (F. F. Chen et al., 2006; DeMars, 2013).

The bifactor model has come into increased popularity and usage in educational research and psychology over the past decade (Reise, 2012). It is a type of hierarchical model with one superordinate (i.e., common factor) and multiple subordinate (i.e., specific) factors (Markon, 2019). The model posits that there is a general factor (i.e., full scale) that accounts for commonalities shared by the facets of a construct and that multiple specific factors (e.g. subscales) account for unique variance in sets of facet-specific items (i.e., scale items) above and beyond the general factor (F. F. Chen et al., 2012). A bifactor model separates the unique variance in item scores accounted for by the specific factors from the common variance shared by the factors (F. F. Chen et al., 2012; DeMars, 2013). The bifactor model is advantageous for psychometric analysis for several reasons. First, the bifactor model allows researchers to determine whether subscales have predictive value over and above the predictive value of the full scale. This information can provide guidance on whether a scale should be treated as unidimensional in analysis, or whether reporting subscale scores and using them in analysis provides any additional benefit beyond that of the full-scale score (DeMars, 2013; Rodriguez et al., 2016b). Second, a bifactor model allows for the testing of measurement invariance of subscale scores in addition to the total scale score, indicating whether subscales measure the same latent construct across groups (F. F. Chen et al., 2006). To our knowledge, despite the benefits of the bifactor model in psychometric evaluation, the model has never been tested for fit for the AMAS for children or adults.

Purpose of the Present Study

The present study aimed to fill the aforementioned gap in the empirical literature by validating the AMAS for use with American middle school populations. Specifically, the present study aimed to examine a) the factor structure and model fit of the scale as indicators of factor validity, b) the equivalence of the factor structure for girls and boys to evaluate factorial invariance across gender, c) the internal consistency of subscale and total scale scores as indicators of reliability, d) correlations of the total scale and subscale scores with measures of theoretically-related constructs as indicators of convergent validity, and e) correlations of the total scale and subscale scores with a measure of a theoretically-unrelated construct as indicators of discriminant validity.

Based on existing literature regarding the reliability and validity of the AMAS in children, adolescents, and adults, the following hypotheses were tested in the sample:

1. A bifactor model will provide an improved fit for AMAS item scores compared to a two-factor solution, with resulting goodness-of-fit indices meeting or exceeding cutoff scores for adequate model fit (CFI $\geq .95$, TLI $\geq .95$, SRMR $\leq .09$, RMSEA $\leq .06$; Hu & Bentler, 1999).

2. The AMAS total scale and MLA and MEA subscales will be invariant across male and female participants, as evidenced by increasingly constrained nested models a) meeting or exceeding cutoff criteria for adequate model fit in both groups (CFI $\geq$ .95, TLI $\geq$ .95, SRMR $\leq$ .09, RMSEA $\leq$ .06; Hu and Bentler, 1999), b) demonstrating decrements in CFI values less than or equal to .01 in magnitude (Cheung & Rensvold, 2002), and c) demonstrating change in RMSEA values less than or equal to .015 in magnitude (F. F. Chen, 2007).
3. The AMAS total scale, MEA subscale, and LMA subscale will demonstrate adequate internal consistency reliability as evidenced by Cronbach alphas and coefficient omegas of .80 or greater (Dunn et al., 2014; Nunnally, 1978; Raykov & Marcoulides, 2011; Rodriguez et al., 2016a, 2016b).
4. The AMAS total scale, MEA subscale, and LMA subscale will demonstrate good convergent validity, as evidenced by large positive correlations ($r \geq .50$) with a measure of math anxiety (e.g., $r = .85$; Hopko et al., 2003) and a measure of test anxiety (e.g., $r = .52$ - .57; Hopko et al., 2003; Primi et al., 2014), medium to large negative correlations ($r \leq -.30$) with a measure of attitudes towards math [e.g., $r = (-.48) - (-.53)$; Primi et al., 2014], medium positive correlations ($.30 \leq r < .50$) with a measure of worry (e.g., $r = .40$; Caviola et al., 2017), and small negative correlations ($r < -.30$) with a measure of interest in STEM careers (e.g., $r = .32$-.38; Huang et al., 2019).
5. The total scale of the AMAS, MEA subscale, and LMA subscale will demonstrate good discriminant validity, as evidenced by smaller correlations with a scale measuring positive affect than with a scale measuring worry.

6. Girls will report significantly more math anxiety than boys, as evidenced by higher mean scores on the AMAS. Mean score differences will represent a small effect, as evidenced by a Cohen's *d* value greater than .20 and less than .50 (e.g., Hedges' *g* = .28; Hill et al., 2016).

CHAPTER THREE

Methods

Participants

As completion of all study materials was not mandatory for participation in the study, participants' responses were included in data analysis if they completed at least 50% of items on the AMAS questionnaire, which was the first questionnaire presented in the study. A total of 604 students from two middle schools in Central and Southeast Texas completed at least 50% of the measure. Of these students, 64 completed study measures in person, while 540 students completed study measures online after their schools closed due to the COVID-19 pandemic.

Measures

Descriptive statistics of questionnaire results are presented in Table B.2.

Abbreviated Math Anxiety Scale (AMAS; Hopko et al., 2003)

The AMAS is a 9-item self-report questionnaire for assessing math anxiety. Respondents rate items on a Likert-type scale ranging from 1 (*low anxiety*) to 5 (*high anxiety*). A total score is derived by summing responses for all items. Total scores range from 1 to 45, with higher scores indicating greater anxiety. As noted above, a modified version of the AMAS has demonstrated coefficient alphas of .86 for the total scale, .80 for the Learning Math Anxiety (LMA) subscale, and .81 for the Math Evaluation Anxiety (MEA) subscale among 11- to 13-year-old children, suggesting good internal consistency for children 11 years of age and older (Carey et al., 2017). The measure has demonstrated good two-week test-retest reliability in adults, with coefficients of .85 for the total scale, .85 for the LMA subscale, and .83 for the MEA subscale (Hopko et al., 2003). A CFA including a modified version of the AMAS suggests that the scale demonstrates good divergent validity from measures of test and general anxiety in middle school children (Carey et al., 2017).

Children's Test Anxiety Scale (CTAS; Wren & Benson, 2004)

The CTAS is a 30-item self-report questionnaire for assessing test anxiety in children. Respondents rate how often they experience thoughts, off-task behaviors, and autonomic reactions related to anxiety on a Likert-type scale ranging from 1 (*almost never*) to 4 (*almost always*). A total score is derived by summing responses to all items. Total scores range from 30 to 120, with higher scores indicating greater anxiety. The CTAS has demonstrated excellent internal consistency with elementary and middle school children ($\alpha = .92$) and good divergent validity with a modified form of the AMAS in children (Carey et al., 2017; Wren & Benson, 2004).

Attitude Toward Mathematics Inventory (ATMI; Tapia, 1996)

The ATMI is a 40-item self-report questionnaire for assessing attitudes toward math. Respondents rate thoughts and feelings towards mathematics and math-related activities on a Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Negatively stated items are reverse scored (i.e., 1 = 5, 2 = 4, 3 = 3, 4 = 2, 5 = 1). Higher scores are associated with more positive attitudes towards math. The ATMI has previously demonstrated moderate to large negative correlations with the AMAS in high school and undergraduate samples [r = (-.48) - (-.53); Primi et al., 2014) and excellent internal consistency in a middle school population (α = .95; Tapia & Marsh, 2000).

Revised Mathematics Anxiety Rating Scale (MARS-R; Plake & Parker, 1982)

The MARS-R is a 24-item self-report questionnaire for assessing math anxiety. Respondents indicate how much anxiety they experience in situations involving mathematics on a Likert-type scale. Response options for each item ranges from 1 (*not at all*) to 5 (*very much)* anxious. Higher scores are associated with more math anxiety. The MARS-R demonstrated a large correlation with the AMAS in a previous study (r = .85; Hopko et al., 2003) and excellent internal consistency among young adults (α = .98; Plake & Parker, 1982).

Penn State Worry Questionnaire for Children (PSWQ-C; Chorpita et al., 1997)

The PSWQ-C is a 14-item self-report questionnaire developed to assess the frequency of worry in children ages 7 to 17 years over and above anxiety and depression. Respondents rate items on a Likert-type scale ranging from 0 (*never true*) to 3 (*always true*). Higher scores are associated with more frequent worrying. The measure demonstrates excellent internal consistency in children ages 5- to 19-years-old (α = .91; Pestle et al., 2008) and has the ability to sufficiently differentiate between children and adolescents who meet criteria for an anxiety disorder and those who do not (Chorpita et al., 1997).

Positive and Negative Affect Schedule for Children, Shortened Version (PANAS-C-SV; Ebesutani et al., 2012)

The PANAS-C-SV is a 10-item self-report measure for identifying anxiety and mood symptoms in children. Respondents rate items on a Likert-type scale ranging from 1 (*very slightly or not at all*) to 5 (*extremely*). Higher scores on the positive affect subscale (PA) and lower scores on the negative affect (NA) subscale indicate fewer mood and anxiety symptoms in children. The PA and NA subscales demonstrate good internal consistency in children ages 6 to 18 years-old, with coefficient alphas of .81 and .86, respectively (Ebesutani et al., 2012). The PA subscale is able to sufficiently discriminate between children and adolescents with and without mood disorders, and the NA subscale is able to sufficiently discriminate between children with and without mood and anxiety disorders (Ebesutani et al., 2012). The PA and NA subscales of the PANAS-C-SV demonstrate larger inter-item correlations than their corresponding scales on the 30-item PANAS-C (Ebesutani et al., 2012). The diagnostic classification accuracy for the PA and NA scales on the PANAS-C-SV do not significantly differ from those on the PANAS-C (Ebesutani et al., 2012).

STEM Career Interest Survey (STEM-CIS; Kier et al., 2014)

The STEM-CIS is a 44-item self-report questionnaire developed to measure interest in pursuing science, technology, engineering, and math classes and careers for middle school students. Respondents rate 11 items on a Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree) for four different vocational field subscales, Science, Technology, Engineering, and Math. Higher subscale scores suggest greater interest in pursuing careers in each of the STEM fields. A total score summing responses from all four subscales suggests respondent interest in the pursuit of a STEM career overall. The items of the STEM-CIS were designed to capture key elements of Lent's (1994) social cognitive career theory, including beliefs about self-efficacy, personal goals, outcome expectations, interest, personal inputs, and contextual supports and barriers (Kier et al., 2014). The Science, Technology, Engineering, and Math subscales demonstrate good internal consistency in students in grades 5-8, with coefficient alphas of .77, .89, .86, and .85, respectively (Kier et al., 2014). The survey also demonstrates good reliability with test-retest correlations of .87 for the full scale and correlations of .67, .73, .89, and .85 for the Science, Technology, Engineering, and Math subscales, respectively (Ünlü et al., 2016). Confirmatory factor analysis of the measure suggests that a four-factor solution provides the best model fit for the items, suggesting that the interpretation of individual subscale scores may provide a better measure of career interest in the four STEM areas than the total scale score (Kier et al., 2014; Ünlü et al., 2016). For the current study, individual subscale scores were interpreted rather than the full-scale score in order to determine the relation between math anxiety and career interest in the four separate STEM categories.

Procedure

Participants were initially recruited in person at a middle school in the Central Texas region. Participants were offered the chance to participate in a research study related to math anxiety. Students received paper packets in math class approximately one week prior to data collection including information about the study and consent and assent forms. Consent forms were offered in both English and Spanish. Students were eligible for participation if they returned a consent form signed by a parent or guardian and an assent form signed by the student. Students who were unable to read and write in English and/or who had an intellectual disability that precluded them from understanding the consent process and/or the administered measures were excluded from recruitment. Eligible participants completed a paper questionnaire packet in their classroom, which included demographic questions and the AMAS, CTAS, ATMI, MARS-R, PSWQ-C, PANAS-C-SV, and STEM-CIS questionnaires. Due to school closures from the COVID-19 pandemic in March 2020 and the associated transition to remote learning via online classes, recruitment and data collection efforts pivoted online. Recruitment extended to a second middle school in Southeast Texas, and all potential participants (students enrolled in math classes at the middle schools) were offered the opportunity to win one of eight $25 gift cards for a major online retailer through completion of the study. Participants provided assent and completed the aforementioned questionnaires through Qualtrics, a study generator program, which was accessible from any device with Internet access. Survey collection was anonymous, with the exception of participants providing the first two letters of their first and last names to allow teachers to disseminate gift cards to the winning students. Gift card winners were chosen at random using a random number generator. All measures and recruitment and data collection practices were reviewed and approved by the Institutional Review Board of Baylor University and the principals of the participating schools.

Data Analytic Strategy

Data analyses were performed using Microsoft© Excel for Mac, Version 16.16.3, IBM SPSS© Statistics© for Mac, Version 26, and R©, Version 3.6.3, using the lavaan and semPlots packages. Responses from the 604 participants who completed at least 50%

of the AMAS were included in confirmatory factor analysis, multigroup confirmatory factor analysis, reliability analysis, and analysis of demographic characteristics. Only participants who completed all study measures were included in correlational analyses for convergent and discriminant validity in order to provide consistent group size across measures.

Missing Data

Questionnaires with at least 50% of items completed were considered to be completed by participants. For survey measures with missing data, a Missing Value Analysis was run to determine whether there were any patterns to the missing responses. Little's test of Missing Completely at Random (MCAR) was also run to determine whether item responses for the questionnaires were missing at random (Little, 1988). For data that was found to be missing at random, item responses were estimated using the Expectation Maximization algorithm, an estimation algorithm for missing data based on maximum-likelihood estimation (Dempster et al., 1977; Schafer & Graham, 2002). For responses that were not found to be missing at random, missing items were not replaced and were considered to represent "0" in data analysis.

Outliers and Normality

Questionnaire responses were assessed for outliers through examination of box-plots of the data. Outliers were defined as data values falling outside of the following ranges: (Q1 – 1.5*IQ) to (3Q +1.5*IQ).[1] Outlier cases were examined and evaluated for exclusion from the data on a case by case basis. Questionnaire responses were examined

[1] Q = Quartile; IQ = Interquartile range

for skewness and kurtosis to determine the normality of the data (Brown, 2014), as violations in the assumption of normality can lead to bias in a number of statistical tests, including correlation, *t*-tests, and analysis of variance (Mishra et al., 2019). Although violation of normality has a lesser impact on data including more than 100 observations, it remains important to correct for non-normal distributions (Mishra et al., 2019). For sample sizes between 50 and 300 participants, *z* scores, computed by dividing the skewness and kurtosis values by their respective standard error, were used to determine normality of the data (Kim, 2013; Mishra et al., 2019). Skewness and kurtosis absolute *z*-values less than 3.29 were considered to fall within the bounds of a normal distribution. For sample sizes larger than 300 participants, histograms and normal Q-Q plots were used to determine normality of the distribution. Absolute skewness and kurtosis values >greater than 2.00 for univariate distributions and absolute values greater than 3.00 for multivariate distributions were considered to indicate non-normality (Bandalos, 2018; Kim, 2013). Data found to be positively skewed was transformed using logarithmic (log) transformation (Keene, 1995).

Model Fit

Confirmatory factor analyses (CFA) were conducted to test the latent structure of the AMAS items (Brown, 2014). Three types of models were tested for fit, 1) a one-factor model with all items loading onto a single latent factor (math anxiety), 2) a two-factor model with items loading onto two correlated latent factors [Learning Math Anxiety (LMA) and Math Evaluation Anxiety (MEA)], and 3) a bifactor model with items loading onto two orthogonal factors (LMA and MEA) and a common g-factor. For the two-factor model, items 1, 3, 6, 7, and 9 of the AMAS were specified to load onto the first factor (LMA) and items 2, 4, 5, 8 were specified to load onto the second factor (MEA). Item coefficients were fixed to zero for the factor that the items were not expected to load on (Bandalos, 2018). The first item loading onto each factor (i.e., item 1 on LMA and item 2 on MEA) were fixed to 1.0 as a marker indicator (Bandalos, 2018; Brown, 2014). For the bifactor model, items were specified to load onto the LMA and MEA factors in the same manner as the two-factor model but were also specified to load onto the g-factor. Given results of the previous EFAs and CFAs of the AMAS highlighted above, a two-factor model was expected to provide a superior fit to the data than a one-factor model. To our knowledge, a bifactor solution has not been previously tested for the AMAS. A bifactor model allows for the simultaneous assessment of both the specific, independent effects of the latent factors (i.e., the LMA and MEA factors) and the common, general effect on the items shared by the factors (i.e., g-factor; F. F. Chen et al., 2012). The bifactor model suggests that the specific factors contribute to effects on the measured items above and beyond that accounted for by the common factor, which accounts for the effects on the items shared among the factors. The factors in the bifactor model are orthogonal, or uncorrelated, with the common g-factor (Chen et al., 2012). For the purposes of this study, the bifactor model would allow us to determine whether the LMA and MEA factors independently contribute to the variance in the AMAS scale items above and beyond their shared common influence (i.e., g-factor). If the bifactor model provided a better fit for the data than the two-factor model, it would provide additional support that the LMA and MEA factors measure separate constructs by controlling for the general factor underlying all item responses (Furtner et al., 2015). Thus, a bifactor model was also tested for data fit.

Maximum likelihood model estimation (ML) was used, as ML tends to produce estimators that are unbiased and consistent even if the model is misspecified (Bandalos, 2018). In ML estimation, the weight matrix is equal to the reproduced covariance matrix of the data (Bandalos, 2018). The weight matrix is updated during each iteration of the CFA process, ensuring that it is fitted to weight the residuals (Bandalos, 2018). The Satorra-Bentler (2010) scaled chi-square and robust standard error adjustments were applied, as violations in the assumption of normality can lead to bias in standard errors and model fit indices (Bandalos, 2018). For non-normally distributed data, the Satorra-Bentler scaled chi-square statistic provides a more accurate indication of model fit than the non-scaled chi-square statistic obtained from Maximum Likelihood (ML) estimation, and Satorra-Bentler robust standard errors also demonstrate greater accuracy than those obtained from ML estimation (Bandalos, 2018).

The chi-square statistic and goodness-of-fit indices were used to evaluate fit of the three models (Bandalos, 2018; Brown, 2014). The chi-square test tests the null hypothesis that the sample covariance matrix is equal to the model-implied covariance matrix (Bandalos, 2018). Therefore, a chi-square test statistic of zero indicates that a model perfectly reproduces the sample covariance matrix (Bandalos, 2018). A significant chi-square value rejects the null hypothesis, suggesting that the model does not provide an adequate fit for the data (Brown, 2014). A non-significant chi-square value that is less than or equal to its degrees of freedom is indicative of a good fitting model (Bandalos, 2018; Brown, 2014; Hooper et al., 2008). However, the chi-square test statistic is highly dependent on sample size and may reject models for discrepancies in fit function that are considered negligible (Bandalos, 2018; Brown, 2014). Because of the tendency for the chi square test to over-reject good-fitting models, the resulting chi-square statistic will not be weighed as highly in assessing model fit for the proposed study as the other fit indices described below. The formula for the chi-square test is $\chi^2 = (n - 1) \times F$, where n is the sample size and F is the minimum value of the fit function (i.e., the difference between the sample matrix and the reproduced covariance matrix; Bandalos, 2018).

The following absolute and comparative fit indices will be used to test model fit based on recommendations by Hu and Bentler (1999): comparative fit index (CFI; Bentler, 1990), Tucker-Lewis index (TLI; Tucker & Lewis, 1973), standardized root mean square residual (SRMR; Bentler, 1995), and the root mean square error of approximation (RMSEA; Steiger & Lind, 1980). Goodness-of-fit index cutoff criteria recommended by Hu and Bentler (1999) will be used to evaluate model fit (CFI $\geq .95$, TLI $\geq .95$, SRMR $\leq .09$, RMSEA $\leq .06$). The results for these indices will be evaluated in conjunction to assess fit for the two-factor model (Hu & Bentler, 1999).

The CFI and TFI indicate the relative improvement in fit of the proposed model over the baseline model, in which all parameters are fixed to zero (Bandalos, 2018). For these indices, values closer to 1.0 suggest improvement in fit over the baseline model (Brown, 2014). Hu and Bentler (1999) proposed a cutoff value of .95 for the CFI and TLI when evaluating model fit. The equation for the CFI is $1 - [(\chi^2_M - df_M) / (\chi^2_B - df_B)]$, where M refers to the factor model and B refers to the baseline model (Bandalos, 2018). The equation for the TLI is $[(\chi^2_B / df_B) - (\chi^2_M / df_M)] / [(\chi^2_B / df_B) - 1]$, where M refers to the factor model and B refers to the baseline model (Tucker & Lewis, 1973).

The RMSEA is the square root of the average of the squared discrepancies between the model's and sample's covariance matrices (Bandalos, 2018). The RMSEA incorporates factorial parsimony into the evaluation of model fit, with values closer to 0 indicating better model fit (Bandalos, 2018). The RMSEA will be equal to 0 if a model's chi-square statistic equals its degrees of freedom, indicating a perfect fit for the model (Bandalos, 2018). The SRMR is a standardized version of the RMSEA, which is computed by turning the sample covariance matrix and model-implied covariance matrix into correlation matrices (Bandalos, 2018). SRMR values closer to 0 suggest fewer discrepancies between the model and sample matrices (Bandalos, 2018). Hu and Bentler (1999) suggest cutoff values of .09 for SRMR and .06 for RMSEA as indicators of a good fitting model. Values at or below these criteria tend to produce lower Type II error rates while minimizing costs of Type I error rates (Hu & Bentler, 1999). The equation for the RMSEA and SRMR are $(\sqrt{\chi^2 - df}) / \sqrt{df\,(n-1)}$ and $\sqrt{\frac{1}{2}\Sigma\left(s_{ij} - I_{ij}\right)^2}$, respectively.

Factorial Invariance

A multigroup confirmatory factor analysis (MGCFA) was used to investigate whether the AMAS demonstrated strong factorial invariance across the sample of girls and boys, in order to allow for the comparison of group means without item-specific biases (Meredith & Teresi, 2006). In order to meet the criteria of strong factorial invariance, all specific factor factors and intercepts had to be equivalent across groups (Meredith & Teresi, 2006). The following sequence of tests were conducted to evaluate configural, metric, and scalar invariance, respectively: 1) simultaneous analysis of equal form, 2) test of equal factor loadings, and 3) invariant intercepts analysis (Brown, 2014; Putnick & Bornstein, 2016). Invariance testing proceeded in a stepwise fashion, in which the least restricted solution was evaluated first, followed by nested models with increasingly restrictive constraints (Brown, 2014; Putnick & Bornstein, 2016).

As a preliminary step, the best fitting model identified in the previous step (the bifactor model) was tested separately in both boys and girls to determine whether it provided an acceptable fit for both groups. CFAs with the bifactor model were conducted using ML estimation with Satorra-Bentler (2010) robust standard error adjustments separately for AMAS item responses for boys and girls. Model fit for both groups was assessed using the CFI, TLI, SRMR, and RMSEA goodness-of-fit indices, with the following cutoffs for adequate model fit (CFI ≥ .95, TLI ≥ .95, SRMR ≤ .09, RMSEA ≤ .06; Hu & Bentler, 1999). Model fit was also assessed using the chi-square statistic, although it carried less weight than the other fit indices due to its tendency to over-reject models with adequate fit (Bandalos, 2018; Brown, 2014). After the bifactor model was found to provide a good fit for both groups, the AMAS scale was tested for configural invariance.

The first step in the MGCFA was to test for configural invariance, or invariance of model form across groups (Putnick & Bornstein, 2016). This step was used to determine whether the subscales of Math Evaluation Anxiety (MEA) and Learning Math Anxiety (LMA) had the same pattern of factor loadings for boys and girls (Putnick & Bornstein, 2016). A CFA with the bifactor model was conducted using ML estimation with Satorra-Bentler (2010) robust standard error adjustments to evaluate model fit independently in boys and girls. Items 3, 6, 7, and 9 of the AMAS were freely estimated by the model for Factor 1 (LMA) and fixed at 0 for Factor 2 (MEA). Item 1 was fixed at 1.0 on Factor 1 as a marker indicator and item 2 was fixed at 1.0 for factor 2 (MEA). For Factor 2, items 4, 5, and 8 were freely estimated by the model, and items 1, 3, 6, 7, and 9 were fixed at 0. All items were freely estimated by the model for the g-factor.

Correlations between the LMA, MEA, and g-factors were set to 0. Model fit for both groups was assessed using the fit indices and chi-square test, as highlighted in the preliminary step.

The second step was to test for metric invariance, or the equivalence of item loadings on the LMA and MEA factors (Putnick & Bornstein, 2016). In order to test for metric invariance, factor loadings were constrained to be equal for both boys and girls (Putnick & Bornstein, 2016). The resulting metric invariance model was then compared with the configural invariance model to assess for discrepancies in fit (Putnick & Bornstein, 2016). Equivalence of model fit was evaluated using Cheung and Rensvold's (2002) and Chen's (2007) suggested criteria of: a) a decrement in the CFI index of .01 or smaller (ΔCFI $\leq$ -.01, and b) a change in the RMSEA index of .015 or smaller (ΔRMSEA $\leq$ .015).

The third step was to test for scalar invariance, which is the equivalence of item intercepts across groups (Putnick & Bornstein, 2016). For this step, the item intercepts (i.e., item means) were constrained to be equal for boys and girls (Putnick & Bornstein, 2016). The resulting scalar invariance model was then compared to the metric invariance model for difference in fit. Change in model fit was again assessed using Cheung and Rensvold's (2002) and Chen's (2007) suggested criteria (ΔCFI $\leq$ -.01 and ΔRMSEA $\leq$ .015).

Internal Consistency

Internal consistency reliability coefficients measure the degree to which items on a scale elicit consistent responses, by measuring correlations between different item composite forms of a scale (Bandalos, 2018). Cronbach's alpha is most commonly used method for assessing internal consistency reliability (Bardhoshi & Erford, 2017; Yang & Green, 2011). The statistic represents the average of all possible split-half reliability coefficients for a scale (Cronbach, 1951). Cronbach's alpha coefficient ranges from 0.0 to 1.0, with values approaching 1.0 suggesting better internal consistency (Yang & Green, 2011). Traditionally, coefficient alpha values.80 have indicated adequate levels of internal consistency for scales in initial development, basic research scales, and scales used for clinical purposes, respectively (Nunnally, 1978; Raykov & Marcoulides, 2011; Yang & Green, 2011). Cronbach's alpha is calculated using the formula $\alpha = (\frac{k}{k-1})(1-\frac{\sum_{i=1}^{k} \sigma_{y_i}^2}{\sigma_x^2})$, where k is the number of items on the test, $\sum_{i=1}^{k} \sigma_{y_i}^2$ is the sum of k item variances, and σ_x^2 is the variance of the total test score (Bandalos, 2018). Because Cronbach's alpha is the most widely used measure of internal consistency in the literature, it is reported in the current study for the AMAS total scale, LMA subscale, and MEA subscale. In line with much of the literature in the social and behavioral sciences, a coefficient alpha of .80 or above was used as a cutoff to indicate adequate internal consistency of the overall scale and subscale.

However, Cronbach's alpha has two major statistical limitations (Rodriguez et al., 2016b; Zwaanswijk et al., 2017). First, Cronbach's alpha does not account for the variance in scores contributed by specific factors independently from common factor variance. Second, the statistic tends to overestimate reliability for multidimensional models because it assumes a unidimensional model. Therefore, for the present study, the coefficient omega (McDonald, 1999), coefficient omega hierarchical (McDonald, 1999), and coefficient omega hierarchical subscale (Reise et al., 2013) will be used to determine reliability of the AMAS total scale and LMA and MEA subscale scores.

Coefficient omega (ω) (McDonald, 1999) is a measure of model-based internal reliability for a multidimensional measure (Reise et al., 2013; Rodriguez et al., 2016a). It indicates the proportion of score variance (in both the full scale and subscales) that can be attributed to both the general and specific factors of the measure, after partitioning out error variance (Hammer & Toland, 2016; Rodriguez et al., 2016a). Coefficient omega is calculated using the following formula,

$$\text{omega} = \frac{\left(\sum_{i=1}^{20} \lambda_{\text{iAlex}}\right)^2 + \left(\sum_{i=1}^{7} \lambda_{\text{iDIF}}\right)^2 + \left(\sum_{i=8}^{12} \lambda_{\text{iDDF}}\right)^2 + \left(\sum_{i=13}^{20} \lambda_{\text{iEOT}}\right)^2}{\left(\sum_{i=1}^{20} \lambda_{\text{iAlex}}\right)^2 + \left(\sum_{i=1}^{7} \lambda_{\text{iDIF}}\right)^2 + \left(\sum_{i=8}^{12} \lambda_{\text{iDDF}}\right)^2 + \left(\sum_{i=13}^{20} \lambda_{\text{iEOT}}\right)^2 + \sum_{i=1}^{20}\left(1 - h_i^2\right)} .$$

Coefficient omega hierarchical (ω_H; McDonald, 1999) indicates the amount of total scale variance attributed to the common factor (g) after partitioning out the variance of specific factors (i.e., LMA and MEA; Hammer & Toldand, 2016; Reise et al., 2013). As such, it indicates the degree to which the total score reflects the common factor (g). Reise et al. (2013) suggest that a $\omega_H > .80$ indicates that the measure predominantly reflects a single common factor, despite its multidimensionality, and that the total score provides a sufficiently reliable measure of the general factor. A coefficient omega hierarchical greater than .80 for the AMAS suggests that the AMAS should be considered unidimensional instrument for the purposes of measurement and that AMAS subscale scores should not be interpreted (Hammer & Toland, 2016). Coefficient omega hierarchical is calculated using the following formula,

$$\text{omegaH} = \frac{\left(\sum_{i=1}^{20} \lambda_{\text{iAlex}}\right)^2}{\left(\sum_{i=1}^{20} \lambda_{\text{iAlex}}\right)^2 + \left(\sum_{i=1}^{7} \lambda_{\text{iDIF}}\right)^2 + \left(\sum_{i=8}^{12} \lambda_{\text{iDDF}}\right)^2 + \left(\sum_{i=13}^{20} \lambda_{\text{iEOT}}\right)^2 + \sum_{i=1}^{20}\left(1 - h_i^2\right)} .$$

Coefficient omega hierarchical subscale (ω_{HS}; Reise et al., 2013) indicates the amount of subscale score variance attributed to the specific factor after accounting for the common factor (g; Hammer & Toland, 2016). A coefficient omega hierarchical subscale value < .50 suggests that the majority of the variance in a subscale's score is due to the general factor, and that the variance due to the specific factor is negligible and does not provide a reliable measure of the intended subconstruct (Hammer & Toland, 2016; Reise et al., 2013). As such, a coefficient omega hierarchical subscale value < .50 for the LMA and MEA subscales suggests that the subscales do not provide reliable measures of learning math anxiety and math evaluation anxiety, and that their scores should not be interpreted (Hammer & Toland, 2016). Coefficient omega hierarchical subscale is calculated using the following formula, omegaHS $= \frac{\left(\sum_{i=1}^{7} \lambda_{\text{iDIF}}\right)^2}{\left(\sum_{i=1}^{7} \lambda_{\text{iAlex}}\right)^2 + \left(\sum_{i=1}^{7} \lambda_{\text{iDIF}}\right)^2 + \sum_{i=1}^{7} \left(1 - h_i^2\right)}$.

Validity

Convergent validity is a type of criterion-related validity, which reflects the degree to which tests that are theoretically expected to measure the same or similar constructs are related (Bandalos, 2018). For the present study, convergent validity was evaluated by examining the Pearson product moment correlations between the total scales of the AMAS with the total scales of the MARS-R, CTAS, ATMI, PSWQ-C, and STEM-CIS. Small, medium, and large effect sizes were defined according to Cohen's (1988) conventions (r = .10, .30, and .50, respectively). As noted previously, both the AMAS and MARS-R were developed for assessing the construct of math anxiety, with the former scale constructed using items from the latter scale (Hopko et al., 2003; Plake & Parker, 1982). Hopko and colleagues (2003) demonstrated a large correlation of r = .85 between the total scales of the AMAS and the MARS-R in an undergraduate sample, suggesting that the AMAS demonstrates good convergent validity with the MARS-R in adults. For the current study, the AMAS total scale was expected to demonstrate a large positive correlation greater than or equal to .85 with the MARS-R total scale in a middle school sample, given that a correlation of this magnitude or greater suggests that the scales are measuring the same construct (Brown, 2014). Large positive correlations were expected between the total scale and subscales of the AMAS and the CTAS, given reported correlations between the total scale of the AMAS and measures of test anxiety ranging from $r = .52$ - $.57$ (Hopko et al., 2003; Primi et al., 2014). Moderate to large negative correlations ($r \leq -.30$) were expected between the total scales of the AMAS and the ATMI, given Primi et al.'s (2014) reported correlations ranging from $r = (-.48)$ - $(-.53)$ between the two measures. Moderate positive correlations were expected between the total scale and subscales of the AMAS and PSWQ-C, given a reported correlation of $r = .40$ between the total scale of the AMAS and a measure of worry in children (Caviola et al., 2017). Moderate negative correlations were expected between the total scale and subscales of the AMAS and the subscales of the STEM-CIS, given Huang et al.'s (2019) reported correlations of $r = -.32$ for boys and $r = -.38$ for girls between measures of math anxiety and career interest in math and science.

Divergent validity assesses the degree to which tests theoretically expected to measure the same or similar constructs are more strongly related than tests expected to measure theoretically different or dissimilar constructs (Bandalos, 2018). For the present study, divergent validity of the AMAS was be assessed by comparing Pearson correlations between the AMAS total scale and subscales and the Positive Affect (PA) scale on the PANAS-C-SV. The PANAS-C-SV is a briefer and more psychometrically sound version of the PANAS-C (Ebesutani et al., 2012). The PANAS-C-SV is therefore expected to produce correlations with measures of worry and general anxiety similar to those of the PANAS-C. The original 15-item PA scale on the PANAS-C demonstrated small to moderate negative correlations with measures of worry and general anxiety in children ranging from r = (-.24) – (-.32; Hughes & Kendall, 2009). Given that the AMAS was developed to measure math anxiety, which is a separate construct from general anxiety (e.g., Suárez-Pellicioni et al., 2016), the AMAS scales are expected to produce smaller correlations with the PA scale than with a scale measuring worry (PSWQ-C).

Mean Differences

Two-tailed independent samples *t*-tests, with the conventional alpha of $p < .05$, were used to examine differences in AMAS total scale and subscale scores by gender. The equation for an independent samples t-test, assuming unequal variances, is $\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$, where $\bar{x}_1$ and $\bar{x}_2$ are the means for the two samples, s_1 and s_1 are the sample standard deviations, and n_1 and n_2 are the sample sizes. For significant results, Cohen's *d* statistic was be computed to determine the size of the effect. Cohen's *d* statistic is an effect size that indicates the difference between two group means measured in standard deviation units (Cohen, 1988). Cohen's *d* is calculated using the formula $d = \bar{x}_1 - \bar{x}_2$, where $\bar{x}_1$ and $\bar{x}_2$ are the group means and $\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$ is the pooled standard deviation of the two groups. The magnitude of the effect was evaluated using Cohen's (1988) conventions, with d = 0.20, 0.50, and 0.80, representing small, medium, and large effects sizes, respectively.

Chi-Square Test of Independence

The chi-square test for independence was used to examine associations between questionnaire completion rates, gender, and school. The chi-square test evaluates the association between two categorical variables. Specifically, the chi-square test evaluates the difference between observed counts and expected counts in the population if the two variables are independent. The formula for the chi-square test is $\chi^2_c = \sum \frac{(O_i - E_i)^2}{E_i}$, where *c* is degrees of freedom, *O* is the observed value, and *E* is the expected value.

CHAPTER FOUR

Results

Descriptive Statistics

Six hundred and four participants completed the AMAS. Of these participants, 422 completed all administered questionnaires (69.9%). Only participants who completed all measures were included in validity correlation analysis. Of survey completers, 74.6% attended a middle school in Central Texas and 25.4% attended a middle school in Southeast Texas. Additional descriptive statistics for the sample are presented in Table B.1.

Study Measures

Descriptive statistics for all study measures are presented in Table B.2.

AMAS. Of the 604 participants who completed the AMAS, two participants produced invalid responses to several items on the scale by either providing more than one response to an item or responding with invalid text response (i.e., responding with the word "anxiety" rather than indicating a Likert response). These responses were excluded, and the affected items were treated as missing data. For the AMAS, less than or equal to 1.0% of item responses were missing. A Missing Value Analysis was run on the AMAS items, with results suggesting that there were no patterns to the missing item responses. Little's test of Missing Completely at Random (MCAR) was run to determine whether item responses on the AMAS scale were missing at random (Little, 1988).

Little's MCAR test was not significant, suggesting that item responses on the AMAS scale were missing completely at random, X^2 (84) = 83.810, p = .485. Through inspection of box plots, the AMAS total scale was found to have one outlier, the LMA subscale was found to have two outliers, and the MEA subscale did not have any outliers. The outliers were kept in the dataset as they appeared to be the product of natural variation in scores rather than aberrant responses. Responses on the AMAS total scale, LMA subscale, and MEA were found to uphold the assumption of normality through examination of histogram, normal Q-Q plot, and skewness and kurtosis values of the data, which were within normal limits.

MARS-R. For the MARS-R, less than or equal to 1.4% of all item responses were missing. A Missing Value Analysis suggested there were no patterns to the missing item responses. Little's test of missing completely at random was significant, suggesting that item responses on the MARS-R were not missing at random, X^2 (346) = 498.82, $p < .001$. Therefore, missing items were not replaced and were represented as "0" in data analysis. One outlier was identified for the MARS-R full scale retained for analysis through examination of a box-plot, as it appeared to be the product of natural variation in scores rather than an aberrant response. The data was normally distributed as assessed by examination of histogram, Q-Q plot, and skewness and kurtosis values of the data. Internal consistency for the scale was good, $\alpha = .96$.

CTAS. For the CTAS, less than or equal to 1.4% of all item responses were missing. A Missing Value Analysis suggested there were no patterns to the missing item responses. Little's test of missing completely at random was significant, suggesting that item responses on the CTAS were not missing at random, X^2 (670) = 824.03, $p < .001$. Therefore, missing items were not replaced and were represented as "0" in data analysis. No outliers were identified through inspection of a box-plot. The data was normally distributed as determined by examination of histogram, Q-Q plot, and skewness and kurtosis values of the data. Internal consistency for the scale was good, α = .95.

ATMI. For the ATMI, less than or equal to 0.9% of all item responses were missing. A Missing Value Analysis suggested there were no patterns to the missing item responses. Little's test of missing completely at random was significant, suggesting that item responses on the ATMI were not missing at random, X^2 (809) = 1205.95, $p < .001$. Therefore, missing items were not replaced and were represented as "0" in data analysis. There were no missing values for the ATMI. Five outliers were identified by box-plot and kept in the dataset, as they appeared to be the product of natural variation in scores rather than aberrant responses. The data was normally distributed as determined by examination of histogram, Q-Q plot, and skewness and kurtosis values of the data. Internal consistency for the scale was good, α = .97.

PSWQ-C. For the PSWQ-C, less than or equal to 0.9% of all item responses were missing. A Missing Value Analysis suggested there were no patterns to the missing item responses. Little's test of missing completely at random was significant, suggesting that item responses on the PSWQ-C were not missing at random, X^2 (150) = 181.29, $p = .042$. One outlier was identified for the PSWQ-C full scale retained for analysis through examination of a box-plot, as it appeared to be the product of natural variation in scores rather than an aberrant response. The data was normally distributed as assessed by examination of histograms, Q-Q plots, and skewness and kurtosis values of the data. Internal consistency for the scale was good, $\alpha = .90$.

STEM-CIS. For the STEM-CIS, less than or equal to 1.4% of all item responses were missing. A Missing Value Analysis suggested there were no patterns to the missing item responses. Little's test of missing completely at random was significant, suggesting that item responses on the STEM-CIS were not missing at random, $X^2\,(1762) = 2024.64$, $p < .001$. Seven outliers were identified by inspection of a box-plot. These responses were kept in the dataset, as they appeared to be the product of natural variation in scores. The data was normally distributed as assessed by examination of histogram, Q-Q plot, and skewness and kurtosis values of the data. Internal consistency for the scale was good, $\alpha = .94$.

PANAS-C-SV. For the PANAS-C-SV positive affect scale, less than or equal to 1.2% of item responses were missing. A Missing Value Analysis suggested there were no patterns to the missing item responses. Little's test of missing completely at random was significant, suggesting that item responses on the PANAS-C-SV positive affect scale were not missing at random, $X^2\,(9) = 14.90$, $p < .05$. Therefore, missing items were not replaced and were represented as "0" in data analysis. No outliers were identified by examination of a box-plot. The data was normally distributed as assessed by examination of histograms, Q-Q plots, and skewness and kurtosis values of the data. No outliers were identified. Internal consistency for the scale was good, $\alpha = .91$.

Group Differences

Descriptive statistics for between group difference tests are presented in Table B.3.

Study Completion

Chi-square independence tests indicated that participants from the middle school in Central Texas were more likely to complete the full survey than participants from the middle school in Southeast Texas, X^2 (1, N = 604) = 24.21, $p < .001$. Completers were also more likely to be female X^2 (1, N = 532) = 9.78, $p = .002$.

Mode of Administration

Independent-samples *t*-tests were performed to determine whether math anxiety ratings differed by mode of questionnaire administration. There was homogeneity of variances, as assessed by Levene's (1960) test for equality of variances for the AMAS total scale ($p = .801$) and the LMA subscale ($p = .067$), but not for the MEA subscale ($p = .008$). Students who completed the questionnaire using paper and pencil had significantly higher scores on the AMAS total scale, $t(602) = 3.40$, $p = .001$, representing a small to medium effect, $d = 0.46$, and MEA subscale $t(602) = 4.50$, $p < .001$ (equal variances not assumed), representing a medium effect, $d = 0.55$, than participants who completed the questionnaire online. There were no significant mean differences between groups on the LMA subscale, $t(602) = 1.80$, $p = .07$.

Gender

Independent-samples *t*-tests were performed to determine whether boys and girls reported different amounts of math anxiety. There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale ($p = .652$), LMA subscale ($p = .710$), and MEA subscale ($p = .515$). Contrary to our hypothesis, there were no significant differences in scores between boys and girls on the AMAS total score, $t(530) = 0.63$, $p = .531$, LMA subscale, $t(530) = 0.06$, $p = .500$, or MEA subscale, $t(530) = 1.13$, $p = .258$.

Race

One-way ANOVAs were performed to determine whether there were group differences in participant reported math anxiety by race. There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale ($p = .775$), LMA subscale ($p = .344$), and MEA subscale ($p = .968$). The mean score differences in math anxiety between racial groups were not significant for the AMAS total score, $F(4, 383) = 1.60$, $p = .174$, LMA subscale, $F(4, 383) = 1.88$, $p = .113$, or MEA subscale, $F(4, 383) = 0.72$, $p = .582$.

Ethnicity

Independent-samples *t*-tests were performed to determine whether students who identified as Hispanic and non-Hispanic reported different amounts of math anxiety. There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale ($p = .576$), LMA subscale ($p = .690$), and MEA subscale ($p = .372$). Hispanic students reported significantly more math anxiety than non-Hispanic students on the AMAS total score, $t(599) = 2.23$, $p = .026$, indicating a small effect, $d = 0.20$. Hispanic students reported significantly more learning math anxiety than non-Hispanic students on the LMA subscale score, $t(599) = 2.42$, $p = .016$, indicating a small effect, $d = 0.21$. However, there were no significant differences in scores on the MEA subscale, $t(599) = 1.49$, $p = .137$.

Parental Level of Education

Highest level of parent education was used as a measure of SES in the sample. One-way ANOVAs were performed to determine whether there were group differences in participant reported math anxiety on the AMAS by parental level of education (proxy for SES). There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale ($p = .457$), LMA subscale ($p = .054$), and MEA subscale ($p = .977$). There were significant differences between parental education groups for the AMAS total score, $F(4, 516) = 2.50$, $p = .042$, and MEA subscale, $F(4, 516) = 2.45$, $p = .046$, but not for the LMA subscale, $F(4, 516) = 1.87$, $p = .115$. However, results of the Tukey-Kramer (1977) post-hoc test did not find any significant differences between groups of parental education levels for AMAS full scale or MEA subscale. The Sidak-Bonferroni (1967) test was run to reduce impact on statistical power and correct for multiple tests, but no significant differences between groups were found for either scale. These contradictory results suggest that the significant omnibus F test findings may be due to type 1 error rather than actual group differences (T. Chen et al., 2018).

Academic Performance

One-way ANOVAs were run to determine whether there was an association between academic performance, as measured by students’ math class grades from the previous semester, and current levels of math anxiety. Math grades were collapsed into three categories due to the small number of students who received grades below B.

Groups were composed of above average performance ($N = 261$), including participants who received A grades, average performance ($N = 180$), including participants who received B grades, and below average performance ($N = 76$), including participants who received C, D, and E/F grades. There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale ($p = .231$), LMA subscale ($p = .712$), and MEA subscale ($p = .078$). The mean differences in math anxiety between academic performance groups were not significant for the AMAS total score, $F(2, 514) = 1.08$, $p = .340$, LMA subscale, $F(2, 514) = 0.26$, $p = .770$, or MEA subscale, $F(2, 514) = 2.06$, $p = .128$.

Mental Health Diagnoses

Attention-Deficit/Hyperactivity Disorder (ADHD). Independent samples *t*-tests were run to determine whether participants with ADHD and participants without ADHD reported different amounts of math anxiety, as measured by the AMAS total score, LMA subscale, and MEA subscale. There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale ($p = .250$), LMA subscale ($p = .211$), and MEA subscale ($p = .563$). There were no significant differences in scores between participants with and without ADHD on the AMAS total score, $t(599) = 0.03$, $p = .976$, LMA subscale, $t(599) = 0.781$, $p = .435$, or MEA subscale, $t(599) = 0.83$, $p = .405$.

Anxiety. Independent samples *t*-tests was run to determine whether participants with anxiety and participants without anxiety reported different amounts of math anxiety, as measured by the AMAS total scale, LMA subscale, and MEA subscale. There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale (p = .070), LMA subscale (p = .419), and MEA subscale (p = .056). There were no significant differences in scores between participants with and without anxiety on the AMAS total score, $t(599) = 1.01$, p = .315 or LMA subscale, $t(599) = 0.21$, p = .832. There was a significant difference in scores on the MEA subscale, $t(599) = 2.16$, p = .031, with participants with anxiety reporting higher levels of math anxiety than participants without anxiety. This represents a small effect $d = 0.29$.

Depression. Independent samples *t*-tests were run to determine whether participants with depression and participants without depression reported different amounts of math anxiety, as measured by the AMAS total scale, LMA subscale, and MEA subscale. There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale (p = .624), LMA subscale (p = .648), and MEA subscale (p = .163). There were no significant differences in scores between participants with and without depression on the AMAS total score, $t(599)$ = .513, p = .608, LMA subscale, $t(599) = 0.526$, p = .599, or MEA subscale, $t(599) = 1.576$, p = .115.

Learning disorders. Of the 12 participants who reported learning disorders, one participant reported a learning disorder in math (8%), nine participants reported learning disorders in reading (75%), and three participants did not qualify their learning disorder (25%). Independent samples *t*-tests were run to determine whether participants with learning disorders and participants without learning disorders reported different amounts of math anxiety, as measured by the AMAS total scale, LMA subscale, and MEA subscale. There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale ($p = .430$), LMA subscale ($p = .442$), and MEA subscale ($p = .228$). Students diagnosed with learning disorders reported significantly more learning math anxiety than students without learning disorders on the AMAS total scale, $t(599) = 2.04$, $p = .041$, indicating a medium effect, $d = 0.53$, and on the MEA subscale, $t(599) = 2.53$, $p = .012$, indicating a small effect, $d = 0.22$. However, there were no significant differences in scores on the MEA subscale, $t(599) = 1.19$, $p = .235$.

Other mental health and neurological disorders. Analyses on the other mental health and neurological disorders surveyed for in the sample (Autism Spectrum Disorder and epilepsy) were not conducted due to the small number of participants who reported these disorders ($N \leq 3$).

Special Education Services

Independent-samples *t*-tests were performed to determine whether participants who have ever received special education services reported different amounts of math anxiety from participants who have never received special education services. There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale ($p = .876$), LMA subscale ($p = .832$), and MEA subscale ($p = .642$). There were no significant differences in scores between students who have received and who have not received special education services on the AMAS total score, $t(601) = 0.20$, $p = .842$, LMA subscale, $t(601) = 0.58$, $p = .561$, or MEA subscale, $t(601) = 0.28$, $p = .777$.

Math Test Recency

One-way ANOVAs were run to determine whether there was an association between the time since participants last took a math test and their amount of math anxiety. There was homogeneity of variances, as assessed by Levene's test for equality of variances for the AMAS total scale ($p = .230$), LMA subscale ($p = .207$), and MEA subscale ($p = .365$). The mean score differences in math anxiety for test recency were not significant for the AMAS total scale, $F(5, 460) = 0.83$, $p = .526$, or LMA subscale, $F(5, 460) = 0.62$, $p = .689$. However, mean score differences were significant for the MEA subscale, $F(5, 460) = 2.48$, $p = .031$. Results of the Tukey-Kramer (1977) post-hoc test indicated that participants who took a math test last week reported significantly more math evaluation anxiety than participants who took a math test more than two weeks ago, $p = .013$.

Model Fit

Fit indices for all models are presented in Table B.4 and differences in model fit are presented in Table B.5. Item loadings are presented in Table B.6, item means are presented in Table B.7, and inter-item correlations are presented in Table B.8.

One-Factor Model

The first model tested was a one-factor (unidimensional) model, in which all items were loaded onto a single latent factor, Math Anxiety. Chi-square results from the ML model estimation with Satorra-Bentler adjustment produced a significant value $X^2(36) = 1871.023$, $p < .001$, which initially suggested that the model did not provide adequate fit for the data. However, due to the strong tendency for the chi-square test to reject good-fitting models due to negligible discrepancies in fit function, other fit indices were weighted more highly in determining model fit (Brown, 2014; Hu & Bentler, 1999). The unidimensional model produced the following fit indices: Robust CFI = .85, Robust TLI = .80, SRMR = .07, Robust RMSEA = .14. When compared to the fit index values recommended by Hu and Bentler (1999) for identifying good-fitting models (CFI ≥ .95, TLI ≥ .95, SRMR ≤ .09, RMSEA ≤ .06), the model did not demonstrate a good fit. Standardized loadings for AMAS factor items ranged from .48 - .72 onto the unidimensional factor. Of note, only one item (Item 3) demonstrated a standardized factor loading with a magnitude above +/- .70, which is commonly suggested as a minimum cutoff point for meaningful item contribution to a scale. The one-factor model is depicted in Figure A.1.

Two-Factor Model

The second model tested was a two-factor (bidimensional) model, in which items were loaded onto two latent factors, Learning Math Anxiety (LMA) and Math Evaluation Anxiety (MEA), based on the findings of previous CFAs of the AMAS (e.g. Hopko, 2003). Chi-square results from the ML model estimation with Satorra-Bentler adjustment produced a significant value X^2 (36) = 81.346, $p < .001$ and the following fit index values: Robust CFI = .96, Robust TLI = .95, SRMR = .04, Robust RMSEA = .07. With the exception of the RMSEA, which was marginally above the cutoff value, comparisons of the model fit index values to Hu and Bentler's (1999) recommended values for model fit suggested that the two-factor model demonstrated a good fit. The two-factor model findings were in line with that of previous research, which found that a two-factor model demonstrated superior fit for the AMAS in children and adults than a single-factor model (e.g., Carey et al., 2017; Caviola et al., 2017; Hopko, 2003). The LMA and MEA latent factors demonstrated a large correlation of $r = .70$, which is similar to previous findings and below the cutoff of .85 for problematic discriminant validity (Brown, 2014). Standardized item loadings on the LMA factor ranged from .49 - .78, with all but two items (items 1 and 9) loading with magnitudes of above .70. Standardized item loadings on the MEA factor ranged from .56 - .78, with all but one item (item 5) loading with magnitudes of above .70. The two-factor model is depicted in Figure A.2.

A chi-square test between the single-factor and two-factor models was significant $X^2(1) = 456.26$, $p < .001$, suggesting a significant difference in fit between the two models. Taking all of these findings into account, the two-factor model was found to demonstrate a good fit for the AMAS and improved fit over the single-factor model.

Bifactor Model

The third model tested was a bifactor model, in which items were fixed and loaded onto the LMA and MEA factors as described for the two-factor model above, with the exception that all items were also set to load freely on the common g-factor. Factors in bifactor models are orthogonal, thus correlations between factors were set to 0. Chi-square results from the ML model estimation with Satorra-Bentler adjustment produced a non-significant value $X^2(18) = 28.66$, $p = .05$, which suggested that the model provided adequate fit for the AMAS items. The bifactor model produced the following fit index values: Robust CFI = .99, Robust TLI = .99, SRMR = .02, Robust RMSEA = .03. Comparisons of the model fit index values to Hu and Bentler's (1999) recommended values for good model fit suggested that the two-factor model demonstrated excellent fit.

Standardized item loadings ranged between .11 - .64 on the LMA factor, .26 - .46 on the MEA factor, and .44 - .68 on the g-factor. The bifactor model is depicted in Figure A.3.

A chi-square test between the bifactor and two-factor models was significant $X^2(8) = 54.51$, $p < .001$, suggesting a significant difference in fit between the two models. Taking the fit results from the single-factor, two-factor, and bifactor models into account, the bifactor model provided an excellent fit and the best fit of the three models tested.

Factorial Invariance

Increasing restricted and nested multigroup confirmatory factor analyses were run using the bifactor model from the previous step to determine whether the factor structure of the AMAS was invariant across groups. Of the 604 participants who completed the AMAS, 72 participants (11.9%) did not report their gender and were excluded from the multigroup confirmatory factor analysis.

Model Fit by Gender

As a preliminary step, the bifactor model, found to provide an excellent fit for the entire sample in the previous step, was evaluated for fit independently in boys and girls to determine whether the model provided a good fit for both groups and if a multigroup confirmatory factor analysis was further indicated. Fit indices for the combined bifactor model and bifactor models for boys and girls only are presented in Table B.9. Responses on all three AMAS scales were normally distributed for boys and girls as determined by examination of histograms, normal Q-Q plots, and skewness and kurtosis values of the data. For the bifactor model for boys and girls combined, chi-square results from the ML model estimation with Satorra-Bentler adjustment produced a significant value, X^2 (18) = 31.93, p = .02, and the following fit index values: Robust CFI = .99, Robust TLI = .98, SRMR = .02, Robust RMSEA = .05, suggesting a good fit for the data. Standardized item loadings ranged between .06 - .69 on the LMA factor, .32 - .42 on the MEA factor, and .44 - .67 on the g-factor.

For the bifactor model using data from girls only, chi-square results from the ML model estimation with Satorra-Bentler adjustment produced a significant value, X^2 (18) = 35.86, p = .007. The bifactor model also produced the following fit index values: Robust CFI = .98, Robust TLI = .96, SRMR = .03, Robust RMSEA = .06, which suggested a good fit for the data. Standardized item loadings ranged between .04 - .71 on the LMA factor, .30 - .45 on the MEA factor, and .48 - .72 on the g-factor. The bifactor model for girls is depicted in Figure A.4.

For the bifactor model using data from boys only, chi-square results from the ML model estimation with Satorra-Bentler adjustment produced a non-significant value, X^2 (18) = 28.29, p = .06. The bifactor model also produced the following fit index values: Robust CFI = .98, Robust TLI = .96, SRMR = .05, Robust RMSEA = .06, suggesting a good fit for the data. Standardized item loadings ranged between .11 - .79 on the LMA factor, .36 - .68 on the MEA factor, and .23 - .83 on the g -actor. The bifactor model for boys is depicted in Figure A.5.

Configural Invariance

Because the bifactor model provided a good fit for both boys and girls independently, a confirmatory factor analysis was run with the two groups nested in order to assess for configural invariance. Fit indices for the nested models are presented in Table B.10. The bifactor model with Satorra-Bentler adjustment using nested data from both groups produced a significant chi-square value, X^2 (36) = 64.05, p = .003 and the following fit index values: Robust CFI = .98, Robust TLI = .96, SRMR = .04, Robust RMSEA = .06. Compared to Hu and Bentler's (1999) recommended fit indices, the model was found to provide a good fit for the data, which supported configural invariance for gender for the measure.

Metric Invariance

Because the measure demonstrated configural invariance for gender, the model was next tested for metric invariance, or the equivalence of item loadings on the factors (Putnick & Bornstein, 2016). Differences in CFI and RMSEA indices were measured to determine whether the model fit for the metric invariance model was significantly different from the model for configural invariance. The model produced a decrease in the CFI index of .003 and an increase in the RMSEA index of .011 from the configural invariance. When compared to Hu and Bentler's (1999) criteria of ΔCFI $\leq$ -.01 and ΔRMSEA $\leq$ .015, item loadings were found to be invariant across groups and metric invariance of the scale for boys and girls was supported.

Scalar Invariance

As metric invariance of gender was supported for the scale, the bifactor model was tested for scalar invariance, or equivalence of item intercepts across groups (Putnick & Bornstein, 2016). For this step, item intercepts (i.e., item means) were constrained to be equal for boys and girls (Putnick & Bornstein, 2016). The model produced an increase in the CFI index of .003 and a decrease in the RMSEA index of .006 from the metric invariance model. Compared to Hu and Bentler's (1999) recommended criteria for change in fit index values, item intercepts were found to be invariant across groups and scalar invariance of the scale for boys and girls was supported.

Reliability

Reliability coefficients are presented in Table B.11.

As measured by Cronbach's alpha, the AMAS full scale demonstrated good internal consistency with a coefficient of 0.86. Also in line with our hypotheses, the Learning Math Anxiety (LMA) Scale and Math Evaluation Anxiety (MEA) Scale demonstrated high levels of internal consistency with Cronbach's alpha of 0.82 and 0.80, respectively.

The coefficient omega (ω) value for the full scale suggested that 89% of variance in the total score was attributable to the common factors (g, LMA, and MEA), after controlling for error variance (Hammer & Toland, 2016). Coefficient omega values for the LMA and MEA subscales suggested that 83% and 81% of variance in the subscale scores were attributed to the common factors of the AMAS, respectively (Hammer & Toland, 2016). These results suggest good internal reliability for the multidimensional total scale and subscales.

The coefficient omega hierarchical (ω_H) value indicated that 74% of the variance in the total score was attributable to the g factor, after partitioning out the variance attributable to the LMA and MEA factors. Although the value was lower than the .80 cutoff proposed by Reise et al. (2013) for determining the unidimensional of scale, the value of .74 is close and suggests that the LMA and MEA subscale scores should be interpreted with caution.

The coefficient omega hierarchical scale (ω_{HS}) values of .26 for the LMA subscale and .20 for the MEA subscale suggest that only 26% and 20% of variance in subscale scores are attributable to the LMA and MEA factors, respectively. These findings also suggest that the majority of the variance in subscale scores is due to the common factor, g. Therefore, the LMA and MEA subscale scores do not reliably measure their intended constructs of learning math anxiety and math evaluation anxiety, respectively. The subscale scores will still be included in analyses for comparison with other studies, as they still account for some of the variance in scores. However, they should be interpreted with caution and the full-scale score should be weighed more heavily for interpretation of results.

Validity

Descriptive statistics for all questionnaires are presented in Table B.2. Correlations between the AMAS scales and all questionnaires are presented in Table B.12. Correlations among validity questionnaires are presenting in B.13.

Convergent Validity

MARS-R. Pearson's product-moment correlations were run to assess the relationships between the total scale of the MARS-R and the AMAS total scale, LMA subscale, and MEA subscale. Inspection of scatter plots suggested linear relationships between the MARS-R and AMAS total scale, LMA subscale, and MEA subscale. In line with our hypotheses, there were statistically significant, large positive correlations between the MARS-R full scale and the AMAS full scale, $r(420) = .70$, $p < .001$, LMA subscale, $r(420) = .59$, $p < .001$, and MEA subscale $r(420) = .68$, $p < .001$.

CTAS. Pearson product-moment correlations were run to assess the relationships between math anxiety, learning math anxiety, math evaluation anxiety, and test anxiety, as measured by the AMAS total scale, LMA subscale, MEA subscale, and CTAS total scale, respectively. Inspection of scatter plots suggested linear relationships between the CTAS total scale and all three AMAS scales. In line with our hypothesis, there were statistically significant, large positive correlations between the CTAS and AMAS total scale, $r(420) = .65$, $p < .001$, CTAS and LMA subscale, $r(420) = .55$, $p < .001$, and CTAS and MEA subscale, $r(420) = .62$, $p < .001$.

PSWQ-C. Pearson's product-moment correlations were run to assess the relationships between the total scale of the PSWQ-C and the AMAS total scale, LMA subscale, and MEA subscale. Inspection of scatter plots suggested linear relationships between the PSWQ-C total scale and all three AMAS scales. As hypothesized, there were statistically significant, medium positive correlations between the PSWQ-C total scale and the AMAS total scale, $r(420) = .47$, $p < .001$, LMA subscale, $r(420) = .36$, $p < .001$, and MEA subscale, $r(420) = .49$, $p < .001$.

ATMI. Pearson's product-moment correlations were run to assess the relationships between the total scale of the ATMI and the AMAS total scale, LMA subscale, and MEA subscale. Inspection of scatter plots suggested linear relationships between the ATMI total scale and all three AMAS scales. Consistent with our hypothesis, there were statistically significant, medium negative correlations between the ATMI and

AMAS full scale, $r(420) = -.46$, $p <.001$, LEA subscale, $r(420) = -.37$, $p <.001$, and MEA subscale, $r(420) = -.45$, $p <.001$.

STEM-CIS. Pearson's product-moment correlations were run to assess the relationships between the Science subscale of the STEM-CIS and the AMAS total scale, LMA subscale, and MEA subscale. Visual inspection of scatterplots suggested that there was not a linear relationship between the Science subscale and the AMAS scales. The lack of significant relationships between the variables was confirmed by non-significant Pearson product-moment correlations with the AMAS full scale, $r(420) = 0$, $p = .865$, LMA subscale, $r(420) = -.06$, $p = .209$, and MEA subscale, $r(420) = .05$, $p = .286$.

Pearson's product-moment correlations were run to assess the relationships between the Technology subscale of the STEM-CIS and the AMAS total scale, LMA subscale, and MEA subscale. Visual inspection of scatterplots suggested that there was not a linear relationship between the Technology subscale and any of the AMAS scales. The lack of significant relationships between the variables was confirmed by non-significant Pearson product-moment correlations with the AMAS full scale, $r(420) = -.04$, $p = .451$, LMA subscale, $r(420) = -.07$, $p = .136$, and MEA subscale, $r(420) = .01$, $p = .825$.

Pearson's product-moment correlations were run to assess the relationships between the Engineering subscale of the STEM-CIS and the AMAS total scale, LMA subscale, and MEA subscale. Visual inspection of scatterplots suggested that there was not a linear relationship between Engineering subscale and any of the AMAS scales. The lack of significant relationships between the variables was confirmed by non-significant

Pearson product-moment correlations with the AMAS full scale, $r(420) = -.05$, $p = .341$, LMA subscale, $r(420) = -.08$, $p = .124$, and MEA subscale, $r(420) = -.01$, $p = .892$.

Pearson's product-moment correlations were run to assess the relationship between the Math subscale of the STEM-CIS and the AMAS total scale, LMA subscale, and MEA subscale. Visual inspection of scatterplots suggested weak linear relationships between the Math subscale and all three AMAS scales. The relationships between the variables were confirmed by statistically significant, small negative correlations between the math subscale and the AMAS total scale, $r(420) = -.21$, $p < .001$, LMA subscale, $r(420) = -.18$, $p < .001$, and MEA subscale, $r(420) = -.20$, $p < .001$.

Pearson's product-moment correlations were run to assess the relationships between the total scale STEM-CIS and the AMAS total scale, LMA subscale, and MEA subscale. Visual inspection of scatterplots suggested that there was not a linear relationship between total scale and any of the AMAS scales. The lack of significant relationships between the variables was confirmed by non-significant Pearson product-moment correlations with the AMAS full scale, $r(420) = -.10$, $p = .050$, $p = .136$, and MEA subscale, $r(420) = -.04$, $p = .374$. However, the STEM-CIS total scale demonstrated a statistically significant, small negative correlation with the LMA subscale, $r(420) = -.12$, $p = .011$.

Divergent Validity

Pearson's product-moment correlations were run to assess the relationship between the positive affect scale of the PANAS-C-SV and the AMAS total scale, LMA subscale, and MEA subscale. Scatter plots of the data suggested weak linear relationships among the scales. As hypothesized, there were statistically significant, negative correlations between the PANAS-C-SV positive affect scale and the AMAS full scale, $r(420) = -.17$, $p < .001$, LMA subscale, $r(420) = -.123$ $p < .001$, and MEA subscale, $r(420) = -.17$, $p < .001$, which were smaller in magnitude than the correlations between the PSWQ-C full scale and the AMAS total scale, $r(420) = .47$, $p < .001$, LMA subscale, $r(420) = .36$, $p < .001$, and MEA subscale, $r(420) = .49$, $p < .001$.

CHAPTER FIVE

Discussion

The purpose of the current study was to evaluate the factor structure and psychometric properties of the Abbreviated Math Anxiety Scale (AMAS) to determine whether it can be a) considered a valid and reliable measure of math anxiety in middle school students, and b) used to compare differences in math anxiety between boys and girls. For our sample of 6-8th grade students from public middle schools in the Southwestern United States, the AMAS demonstrated good internal consistency and strong convergent and divergent validity with a number of established scales. A bifactor model provided an improved fit for the AMAS over one-factor and two-factor models, and the bifactor model was invariant across gender at the configural, scalar, and metric levels.

Factor Structure

In order to confirm the factor structure of the AMAS and assess model fit in a middle school population, we performed confirmatory factor analyses for one-factor, two-factor, and bifactor models. Previous studies have found the two-factor model to provide a good fit for the AMAS in college and elementary school populations in the United States and abroad (Carey et al., 2017; Caviola et al., 2017). We therefore expected a two-factor model to provide an improved fit for the data compared to a unidimensional model, with items loading onto two subscales, Learning Math Anxiety (LMA) and Math Evaluation Anxiety (MEA; Hopko et al., 2003). We also sought to determine whether a bifactor model would provide a good and improved fit for the data compared to a two-factor model, as bifactor models allow for the simultaneous assessment of the common and independent effect of specific, latent factors on scale items, and are increasingly being used in the field of psychological research (Bornovalova et al., 2020). In line with our hypotheses, we found that the two-factor model provided a good fit for the data and an improved fit over the one-factor model, which did not provide an adequate fit. Further, we found that the bifactor model provided a superior fit for the data over the two-factor model, suggesting that the specific factors of LMA and MEA independently contribute to the variance in AMAS item responses above and beyond their shared common influence, or the g-factor. These findings suggest that the LMA and MEA subscales independently provide information about participant responses above and beyond the full-scale and should be considered in future use of the AMAS as a research and clinical tool.

Of note, two items produced small item loadings across all models tested. Specifically, item 1 "Having to use tables in the back of a math book" and item 5 "Being given a homework assignment of many difficult problems that is due the next class meeting" produced standardized item loadings of less than .60 for the one-factor and two-factor models. These findings suggest that item 1 and item 5 may not be particularly salient items for assessing math anxiety in this population. Closer examination of item means presented in Table B.7 suggests that participants reported relatively lower scores for item 1 and relatively higher scores for item 5 compared to other items. However, examination of the inter-item coefficients presented in Table B.8 indicates that the magnitude of the correlations between items 1 and 5 and the other items on their respective subscales were large enough to indicate that they measure similar constructs (e.g. r >.20; Piedmont, 2014). We propose several explanations for the weak loadings of these items onto the full scale and subscales of the AMAS.

First, as schools are increasingly integrating technology into the classroom, especially during the COVID-19 pandemic, it is possible that many students have adopted the use of digital math texts rather than printed books for learning and completing assignments. It is, therefore, possible that today's school-age students are not familiar with the process of turning to the back of a math book to check for reference tables. This hypothesis is supported by the author's anecdotal experience in administering the AMAS questionnaire in-person, when several students expressed confusion regarding the meaning of item 1 and indicated that they did not use printed textbooks in math class. Due to changes in the modern learning environment since the AMAS was first published in 2003, item 1 may no longer capture the construct that it was intended to measure.

Second, compared to the other items on the MEA scale that measure anxiety related to tests and quizzes, item 5 pertains to anxiety related to completing homework. Because homework may be viewed as a less threatening task than quizzes or tests, it is possible that item 5 measures a closely related, yet distinct construct (e.g. Math Assignment Anxiety). However, previous exploratory factor analyses of the items in other samples have found item 5 to load adequately onto the MEA subscale. Thus, we hypothesize that item 5 measures a subsection of evaluation math anxiety that participants in this particular sample found less anxiety provoking.

Third, it is possible that both items represent additional latent factors that are not accounted for by the model. Given that the two items do not correlate strongly (r = .26), it is unlikely that they would load onto the same latent factor and would represent separate factors. However, taking into account results of previous exploratory and confirmatory analyses supporting a two-factor solution for the AMAS, as well as the principle of parsimony (Vandekerckhove et al., 2015), this explanation is unlikely.

Factorial Invariance

The bifactor model was found to provide a good fit for both boys and girls in the sample. Results of a multigroup confirmatory factor analysis indicated that the model was equivalent for boys and girls across form, factor loadings, and intercepts. Because configural, metric, and scalar invariance was supported between boys and girls, these results suggest that the AMAS demonstrates strong factorial invariance across gender for middle school students and can be used to measure and compare mean differences in math anxiety in middle-school aged boys and girls in an unbiased manner.

Reliability

In line with our hypothesis, the full scale, LMA subscale, and MEA subscales of the AMAS demonstrated high levels of internal consistency reliability ($\alpha \geq .80$), indicating that participants responded to the different items on the full scale and subscales in a similar manner. These results support the reliability of the measure.

Although Cronbach's alpha is the most commonly used statistic for measuring internal consistency in the measurement literature, it assumes a unidimensional measure. Therefore, we also measured coefficient omega, coefficient omega hierarchical, and coefficient omega hierarchical subscale to account for the multidimensionality of the measure and control for variance attributable in scores to error. Our findings suggested that 89% of variance in the total score, 83% of variance in LMA scores, and 81% of variance in MEA scores were attributable to the common factor and specific factors after controlling for error variance. These results suggest that the total scales and subscales of the AMAS demonstrate good internal reliability. The coefficient omega hierarchical value indicated that the majority of variance in the total scale (74%) was attributable to the common factor rather than the LMA and MEA factors. Reise et al. (2013) suggest that value greater than or equal to 80% indicate that a measure should be treated as unidimensional for the purpose of interpretation, despite its multidimensionality. Although the coefficient hierarchical value did not fall above this cutoff for the AMAS, the finding suggests that the full scale should be weighed more heavily in interpretation than the LMA and MEA subscales, which do not appear to provide a substantial increase in value above the total scale. Specifically, the coefficient omega hierarchical subscale values for the LMA and MEA subscales suggest that close to only one quarter of the variance in items is due to these factors alone, independent of their shared contribution to the variance. In accordance with Reise and colleagues (2013), the results suggest that the subscales do not provide reliable measures of their intended constructs independent of the total scale. Thus, the scores of the LMA and MEA subscale scores should be interpreted with caution relative to the full scale, as the total scale accounts for the majority of variance in item responses. However, because the coefficient omega hierarchical was less than the suggested cutoff for unidimensionality, we have also included the LMA and MEA subscales in our analyses in addition to the total scale. However, we recommend that the full scale be the primary measure of math anxiety interpreted for clinical and research purposes.

Of note, the present study initially aimed to collect AMAS responses at two different time points in order to assess the test-retest reliability of the measure. However, school closures and the necessary pivot to online data collection due to the COVID-19 pandemic prevented us from collecting data at a second time point. Specifically, we chose not to collect identifying information through online data collection in order to ensure the privacy of our minor participants and the anonymity of responses. Although such information would have allowed us to match participant responses across time points, we did not believe that it would support our participants' best interests to do so given their age and the sensitive nature of the mental health data collected.

Validity

Convergent and Discriminant Validity

The AMAS full scale, LMA subscale, and MEA subscale demonstrated large, statistically significant correlations with the total scale of the MARS-R, a validated scale for the assessment of math anxiety in college students. Although the magnitude of the correlations did not reach a cutoff proposed by Brown (2014) for demonstrating that two scales measure the same construct ($r \geq .85$), the AMAS full scale demonstrated a strong correlation with the full scale of the MARS-R ($r = .71$). We therefore argue that the strength of the association between the two scales suggests that they measure similar constructs. Moreover, as the MARS-R contains items that are intended for college students rather than middle school students, it is possible that the differences in participants' responses to the two measures were due to the relevance of item content for this sample. The magnitude of the correlations between the LMA and MEA subscales and the full scale of the MARS-R were not as strong as the correlation with the full scale of the AMAS. This was not entirely surprising, given that the subscales are intended to measure facets of math anxiety, whereas the AMAS and MARS-R full scales are intended to measure the whole construct. However, given that the correlations of the subscales with the MARS-R full scale were not particularly strong ($r \leq .62$), these findings may suggest that the AMAS full scale is a better measure of math anxiety than the individual subscales.

In line with our hypotheses, the AMAS full scale, LMA subscale, and MEA subscale produced large positive correlations with a measure of test anxiety, medium positive correlations with a measure of worry, moderate negative correlations with a measure of attitudes towards math, and small negative correlations with a measure of math career interest. The AMAS scales also produced smaller correlations with a measure of positive affect than a measure of worry. These findings suggest that the AMAS full scale and subscales demonstrate good convergent and discriminant validity, which further supports the validity of the measure in this population.

Surprisingly, while math career interest was significantly negatively correlated with the AMAS scales, there was no significant relationship between the AMAS scales and career interest in science, technology, or engineering. It is possible that with the rising interest, visibility, and availability of careers involving science, technology, and engineering, math anxiety in middle school may not be as strong of a predictor of career interest in these fields as previously suggested in the literature (e.g., Ahmed, 2018). Given that avoidance of STEM-related career paths is more strongly associated with math anxiety in high school-aged youth, these findings may point to middle school as an ideal time for introducing interventions to reduce the impact of math anxiety on the pursuit of higher learning in math and related fields among youth and young adults.

Demographic Characteristics

Gender

Contrary to our hypothesis, there were no significant mean differences in AMAS full scale or subscale scores between male and female participants. This finding was unexpected, given the large body of research documenting higher reported levels of math anxiety among girls than boys (e.g., Bieg et al., 2015; Dowker et al., 2016; Hill et al., 2016). Several possibilities are suggested to account for this unexpecting finding.

The mean score for both genders suggests that boys and girls did not report particularly high levels of math anxiety overall. On average, boys and girls provided Likert response scores of 2.42 and 2.47 per item, respectively, with item responses ranging from 1 (Not at All) to 5 (Very Much) anxiety. As state tests were cancelled for the majority of participants during the COVID-19 pandemic and many schools substantially reduced student academic demands and expectations during this time (Kuhfeld et al., 2020), it is possible that both boys and girls experienced lower levels of math anxiety overall. It is also possible that participant math anxiety was overshadowed by the substantial and numerous stressors impacting youth during the pandemic, including fears related to the health and safety of friends and family members, financial concerns, worries about the future, and the impact of social isolation on mood (Golberstein et al., 2020). Although data is still limited on the impact on child and adolescent mental health, preliminary research suggests that the pandemic and associated lockdowns are associated with increased anxiety, depression, substance use, and stress in youth, as well as possible increases in child abuse and neglect (Jones et al., 2021; Marques de Miranda et al., 2020; Ortiz et al., 2021). Therefore, anxiety associated with math learning and evaluation may not have been as salient relative to other stressors.

It is also possible that the schools sampled for this study provide support for students in a manner that reduces gender-specific factors impacting math anxiety, or that the gender discrepancy in math anxiety is shrinking in Texas, or possibly nationwide. If the lack of measured difference in math anxiety between boys and girls in this study is due to a local or national trend, it would provide support for future equality of opportunity and achievement in STEM between men and women.

Race, Ethnicity, and Socioeconomic Status

In line with previous findings in the literature, there were no significant mean differences in AMAS full scale or subscale scores among participants in our sample from different racial or socioeconomic groups (e.g. Hart & Ganley; Hembree, 1990). However, there was a small effect of ethnicity on math anxiety, with Hispanic students reporting higher levels of math anxiety than non-Hispanic students on the full scale and LMA subscale. These results suggest that Hispanic students may especially benefit from interventions targeting math anxiety. These findings are particularly salient in light of the discrepancy in STEM-employment rates for Hispanic individuals relative to individuals of White non-Hispanic and Asian heritage (Funk & Parker, 2018). In line with the recommendations of Fernández and colleagues (2021), future studies should investigate whether interventions targeting study skills in Hispanic students have beneficial effects on reducing math anxiety within this population.

Math Performance and Math Test Recency

Our results did not demonstrate mean differences in math anxiety between participants who were high and low achieving in math, as measured by their grades. This finding was unexpected due to the negative relationship between math anxiety and math performance extensively described in the literature (e.g., Devine et al., 2012; Ma, 1999). The results from our sample may be attributable to participant grades providing too broad a measure of math performance, as 50% of participants reporting receiving A grades in their math course last semester. Thus, the measure of math achievement included in this study may not provide a valid measure of the construct, providing and explanation for the discrepancy between our findings and those of previous studies.

We found that students who most recently took a math test a week before participating in the study reported more math evaluation anxiety than students who most recently took a math test more than two weeks before participating. No other group differences were found related to time since most recent math test for the MEA subscale, LMA subscale, or AMAS full scale. The overall lack of group differences is consistent with Conlon and colleagues' (2021) finding that ratings of math anxiety did not differ by time point, with the exception of directly following a math test, and support the conceptualization of math anxiety as a state rather than trait form of anxiety (e.g., Ashcraft & Ridley, 2005; Bieg et al., 2015).

Mental Health Diagnoses

Although participants with anxiety reported higher levels of math evaluation anxiety on the MEA scale than participants without anxiety, there were no differences in scores on the total scale or LMA subscale. These findings are consistent with the evaluation of a student's performance, rather than the act of learning the information, being the driving force of anxiety in these participants. Conversely, participants with learning disorders reported more math anxiety on the total scale and learning math anxiety on the LMA subscale, but not on the MEA subscale, than their peers. This finding is consistent with participants with learning difficulties (in math and/or other academic areas) reporting higher levels of anxiety related to the act of learning math, and is in line with expected results. Surprisingly, participants who have previously received special education services did not report differing amounts of math anxiety on the full scale, LMA subscale, or MEA subscale than participants who had never received special education services. These findings may be due to the very small percentage of the sample who ever received special education services (7.1%) and/or the wide range of special education services that participants received. Although, a previous study found that college students with ADHD reported significantly more math anxiety than their peers (Canu et al., 2007), we did not find group differences in math anxiety for students with and without ADHD in our sample. Once again, our findings were likely impacted by the very small percentage of our students who reported diagnoses of ADHD (10.8%) in our sample.

Limitations

This study has a number of limitations. First, the majority of data was collected several months into nationwide lockdowns imposed in response to the COVID-19 pandemic. It is possible that changes in participant schedules, learning environments, and levels of stress related to environmental factors (e.g., COVID-19 related worry, economic impact, isolation, increased mental health difficulties) may have impacted ratings of math anxiety and other constructs.

Second, data was collected at two separate time points, with some students participating in the school setting with questionnaires administered via paper-and-pencil, while other students participated at home with questionnaires administered electronically. Although, research suggests that response differences to questionnaires administered electronically versus on paper tend to be negligible (Gwaltney et al., 2008; Mangunkusumo et al., 2005; Muehlhausen et al., 2015), our results indicated that participants who completed the questionnaire electronically reported significantly less math anxiety than participants who completed the survey on paper. Although it is unclear whether these differences in scores are due to demand characteristics, setting, modality, effects of the pandemic, or other factors, these differences may have impacted our overall findings.

Third, as we collected data from school children under the age of majority, which is a specially protected population, we did not penalize participants for skipping items that they did not wish to respond to. As a result, demographic information is missing for a number of participants and approximately 30% of participants who began the survey did not complete all measures. Notably, our sample's rate of completion was very close to those of large scale online surveys with youth participants (Anderson & Jiang, 2018; Larson et al., 2011), suggesting that our completion rate was within expected limits.

Fourth, much of the data was gathered in the last months of the school year after students had completed much of their math testing and learning. Thus, the timing of data collection may have impacted participants' reported levels of math anxiety.

Fifth, our sample population was strongly homogenous in that the large majority of participants in the sample were White, not Hispanic, in 7th grade, and had no mental health diagnoses. Therefore, results of the study may not generalize to other populations and the study would benefit from replication in a more diverse sample.

Sixth, we were unable to collect AMAS data from participants at different time points, as initially proposed for the study, due to the pivot from in-person data collection to online data collection during the pandemic. As a result, we were unable to determine test-retest reliability for the AMAS full scale and subscales in this sample. Although the AMAS has been found to demonstrate good test-retest reliability in adult populations, future research should seek to measure test-retest reliability in middle school-aged students as well.

Future Directions for Research

As the bifactor model provided an improved fit for the AMAS over the two-factor model previously identified for adult and child samples, we suggest that the bifactor model should be tested with other populations to determine whether it provides improved fit across populations. We also suggest that further research continue to explore the relative advantage of using the full scale of the measure relative to the subscales in different populations.

Due to the relative homogeneity of the sample for the current study, this study should be replicated in a more ethnically, racially, and socioeconomically diverse school sample in order to determine the generalizability of results across populations. The current study would also benefit from replication following the COVID-19 pandemic, to determine whether changes due to lockdowns and quarantine impacted participant responses.

As we did not find differences in mean AMAS scores between boys and girls in this sample, we recommend that the measure be further tested for invariance in a sample that better reflects gender differences in math anxiety documented in the literature (e.g., Dowker et al., 2016). Because of the increasing integration of electronic tools and the Internet in the classroom, the AMAS would likely also benefit from an update to the terminology and content of items to account for changes in the learning environment.

Clinical Implications

Studies suggest that a wide range of clinical interventions may be effective for reducing math anxiety in youth. Interventions include sustained exposure to math cues, scaffolding parent-child interactions involving math, expressive writing, breathing, mindfulness, and positive affirmation exercises, and changing cognitive appraisals of math and physiological cues of anxiety (Brunyé et al., 2013; Jamieson et al., 2016; Luttenberger et al., 2018; Park et al., 2014; Ramirez et al., 2018; Samuel & Warner, 2019; Supekar et al., 2015). Results of these studies support the implementation of what are likely low-cost, short-term, psychological treatments for math anxiety that can be implemented in home, school, and/or therapy environments.

Despite the limitations of the current study, our results support the use of the AMAS as an outcome measure for clinical practice and research focusing on reducing math anxiety in middle school-aged populations. The robust psychometric properties of the measure, including its factor structure, internal consistency and convergent and discriminant validity suggest that the AMAS demonstrates strong construct validity for math anxiety in this population. Because the measure demonstrates factorial invariance across gender, it can be readily used to accurately compare boys' and girls' levels of math anxiety at baseline, during, and following clinical intervention. The short length of the AMAS also makes it an ideal measure for use in clinical work and research with school-age populations, as it is not cumbersome to administer, respond to, or score. The findings of the current study highlight middle school as an ideal time for implementing interventions for reducing math anxiety, as well as support the use of the AMAS as a reliable and valid scale for measuring the outcomes.

Conclusion

In a homogenous community sample of middle school students, the AMAS was found to be a valid and reliable measure of math anxiety. The AMAS demonstrated factorial invariance for gender, suggesting that the items are interpreted in a similar manner by both boys and girls. A bifactor model provided the best fit for the measure, suggesting that the Math Evaluation Anxiety and Learning Math Anxiety subscales contribute additional variance in scores over and above the total scale. Overall, our findings suggest that the AMAS is a psychometrically sound measure of math anxiety for use in middle school-aged populations with similar demographic characteristics to our sample and can be used to compare differences in math anxiety between boys and girls in an unbiased manner.

APPENDICES

APPENDIX A

Figures

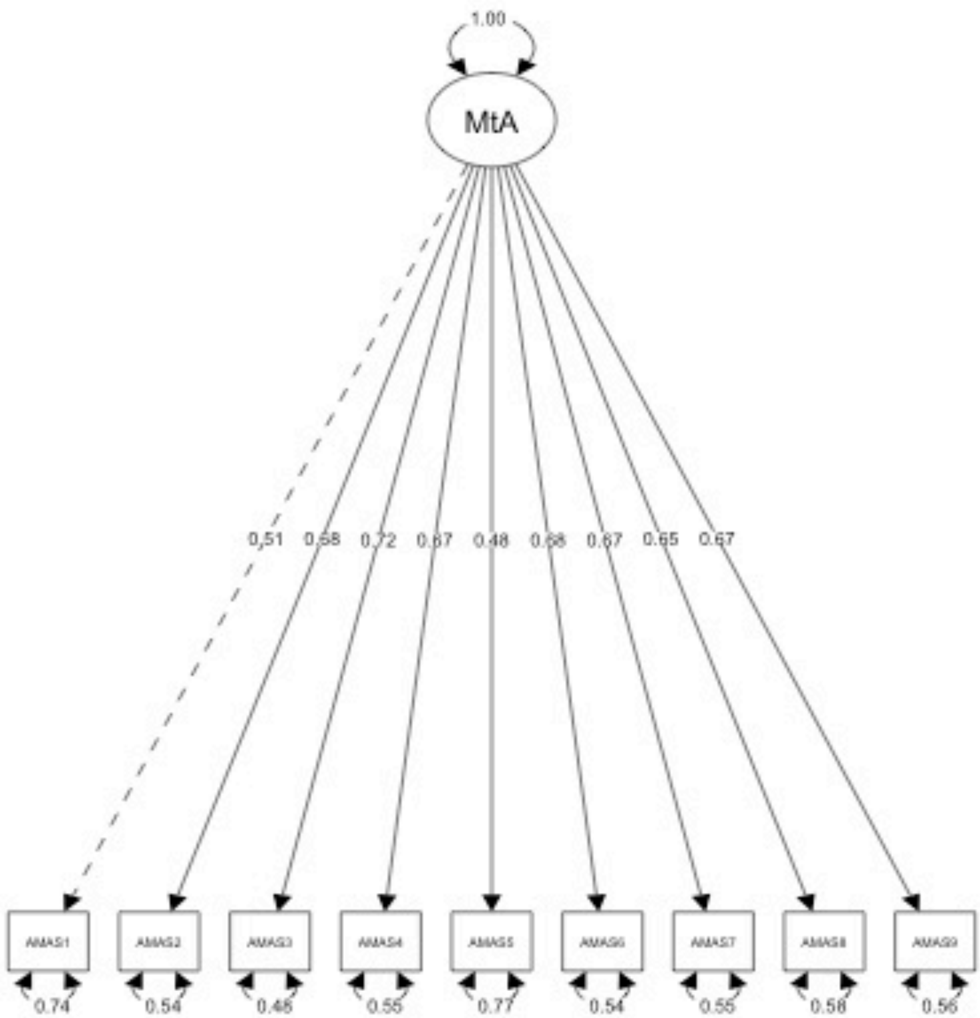

Figure A.1. One-Factor Model of the AMAS

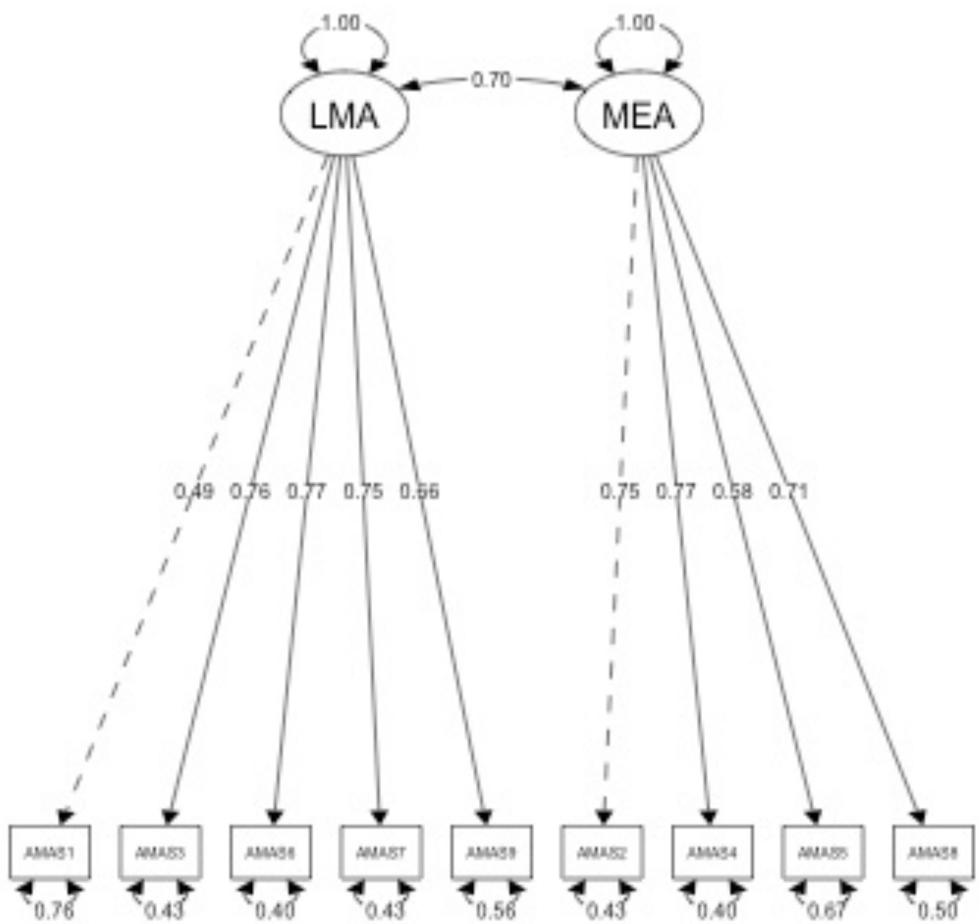

Figure A.2. Two-Factor Model of the AMAS

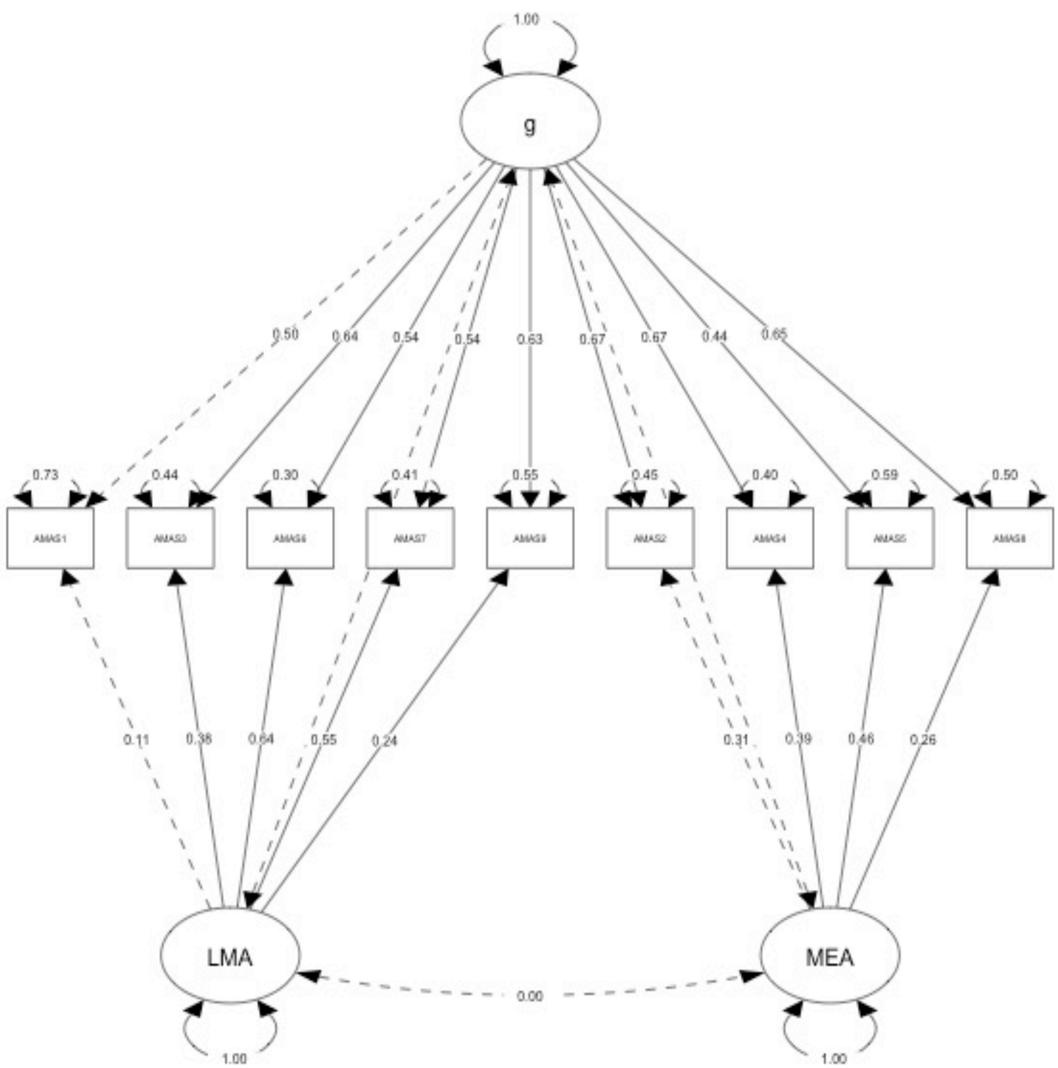

Figure A.3. Bifactor Model of the AMAS

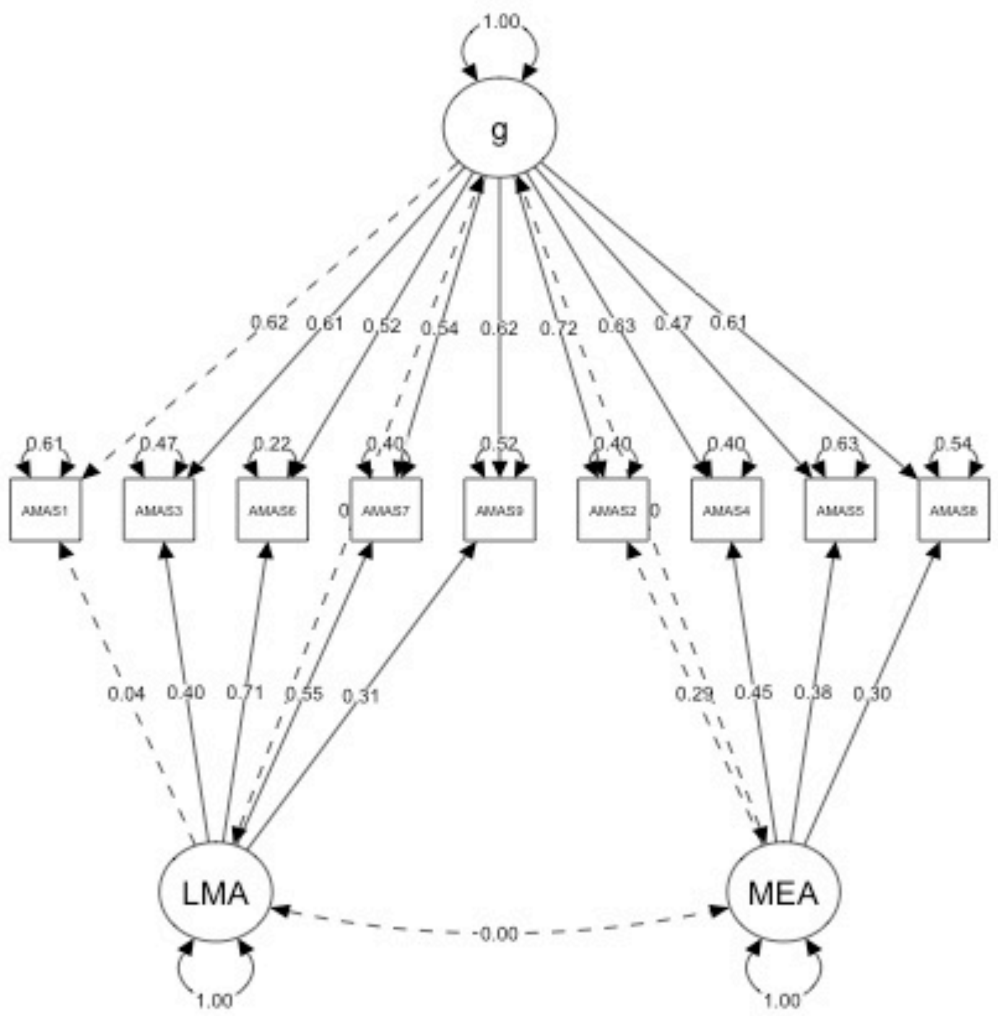

Figure A.4. Bifactor Model of the AMAS for Girls

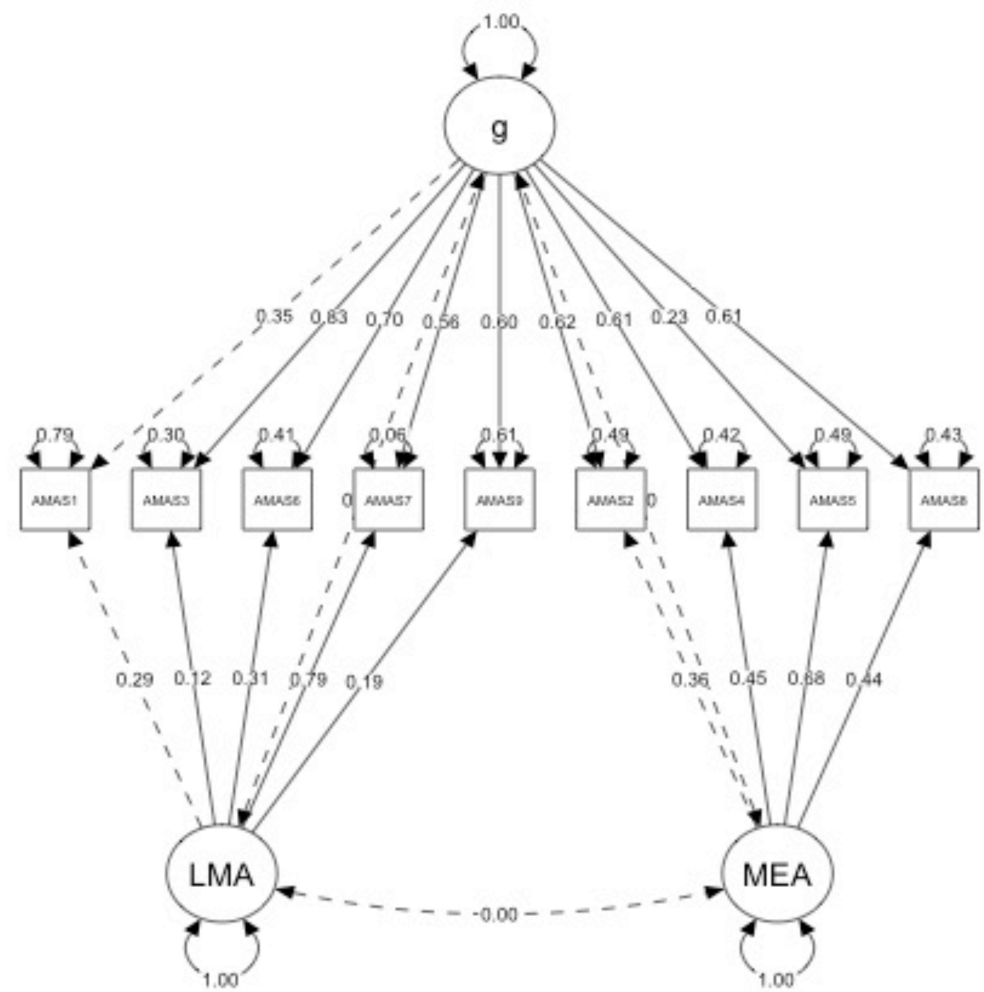

Figure A.5. Bifactor Model of the AMAS for Boys

APPENDIX B

Tables

Table B.1

Sample Characteristics

Characteristic	*n*	Percent of Sample
Gender		
Male	203	33.6
Female	329	54.5
Missing	72	11.9
Grade		
5	1	0.2
6	39	6.5
7	330	54.6
8	158	26.2
School Location		
Central TX	413	68.4
Southeast TX	191	31.6
Race		
White	275	45.5
Black	42	7.0
Asian	12	2.0
Other	16	2.6
Multiracial	43	7.1
Missing	216	35.7
Ethnicity		
Hispanic	198	32.8
Not Hispanic	403	66.7
Missing	3	0.5
Parental Level of Education		
Some High School	63	10.4
High School or GED	72	11.9

(Continued)

Characteristic	*n*	Percent of Sample
Some College or Vocational School	63	10.4
College	201	33.3
Graduate Degree	122	20.2
Missing	83	13.7
Math Grades		
A	261	43.2
B	180	29.8
C	58	9.6
D	14	2.3
E/F	4	0.7
Missing	87	14.4
Mental Health Diagnoses		
ADHD	65	10.8
Anxiety	54	8.9
Depression	28	4.6
Autism Spectrum Disorder	3	0.5
Learning Disorder	12	2.0
Epilepsy	2	0.3
Special Education Services		
Ever Received Special Education Services	43	7.1
Never Received Special Education Services	560	92.7
Missing	1	0.2
Math Test Recency		
Today	95	15.7
Yesterday	56	9.3
This Week	56	9.3
Last Week	140	23.2
Two Weeks Ago	23	3.8
More than Two Weeks Ago	96	15.9
Missing	138	22.8

Note. $N = 604$. Participant age ranged from 10-15 years-old with a mean age of 12.99 years ($SD = .78$).

Table B.2

Descriptive Statistics for Questionnaires

Instrument	*N*	Mean	SD	Skewness Statistic	Skewness SE	Kurtosis Statistic	Kurtosis SE
AMAS Full Scale	604	22.04	7.55	.33	.10	-.67	.20
LMA Scale	604	10.48	4.52	.64	.10	-.43	.20
MEA Scale	604	11.56	3.98	.28	.10	-.62	.20
RMARS	422	50.95	20.20	.60	.12	-.36	.24
CTAS	422	66.19	19.73	.28	.12	-.48	.24
ATMI	422	136.23	30.10	-.21	.12	-.04	.24
PSWQ-C	422	34.41	10.28	.35	.12	-.51	.24
STEM-CIS Science Scale	422	40.05	7.75	-.33	.12	.18	.24
STEM-CIS Technology Scale	422	40.00	8.83	-.41	.12	.38	.24
STEM-CIS Engineering Scale	422	35.93	9.47	-.05	.12	.03	.24
STEM-CIS Math Scale	422	39.13	8.10	-.41	.12	.18	.24
STEM-CIS Total Scale	422	155.10	26.66	-.28	.12	.53	.24
PANAS-C-SV Positive Affect	422	15.64	4.99	.07	.12	-.57	.24

Table B.3

Descriptive Statistics for Between Group Difference Tests

Demographic Characteristic		AMAS Full Scale		LMA Subscale		MEA Subscale	
Mode of Administration	*N*	Mean	SD	Mean	SD	Mean	SD
Paper	64	25.08	7.44	11.43	4.19	13.65	4.51
Electronic	540	21.68	7.44	10.36	4.55	11.32	3.84
Gender							
Boys	203	21.76	7.59	10.43	4.64	11.33	3.86
Girls	329	22.19	7.64	10.46	4.58	11.73	4.02
Race							
White	275	21.53	7.82	10.09	4.79	11.44	4.03
Black	42	23.79	7.51	11.60	4.39	12.19	3.86
Asian	12	18.00	6.54	7.83	3.19	10.17	3.69
Other	16	21.88	8.49	10.31	5.17	11.56	3.97
Multiracial	43	20.88	7.25	9.67	4.20	11.20	4.00
Ethnicity							
Hispanic	198	23.00	7.66	11.11	4.53	11.89	3.80
Not Hispanic	403	21.54	7.30	10.16	4.51	11.38	4.06
Parental Educational Level							
Some High School	63	21.39	7.27	10.33	4.28	11.06	4.04
High School or GED	72	23.15	7.28	11.11	4.29	12.04	3.76

(Continued).

Demographic Characteristic		AMAS Full Scale		LMA Subscale		MEA Subscale	
	N	Mean	SD	Mean	SD	Mean	SD
Some College or Vocational School	63	23.40	8.15	11.24	5.12	12.15	3.89
College	201	22.24	7.56	10.36	4.60	11.87	4.02
Graduate Degree	122	20.38	7.54	9.63	4.44	10.75	3.96
Math Grades							
Above Average (A)	261	21.65	7.49	10.28	4.50	11.37	3.97
Average (B)	180	21.88	7.47	10.41	4.62	11.46	3.76
Below Average (C, D, E/F)	76	23.11	8.73	10.71	5.05	12.40	4.39
Mental Health							
Diagnoses							
ADHD	65	22.00	6.94	10.06	4.25	11.94	3.76
No ADHD	536	22.03	7.65	10.53	4.57	11.50	4.00
Anxiety	54	23.01	8.20	10.35	4.71	12.66	4.35
No Anxiety	547	21.93	7.50	10.49	4.52	11.44	3.93
Depression	28	22.74	7.80	10.04	4.43	12.71	4.40
No Depression	573	21.99	7.56	10.50	4.54	11.49	3.95
Learning Disorder	12	26.43	9.45	12.02	6.07	14.42	4.64
No Learning Disorder	589	21.94	7.51	10.44	4.50	11.49	3.95
Special Education							
Ever Received	43	21.81	7.83	10.09	4.89	11.72	4.17
Never Received	560	22.05	7.54	10.51	4.50	11.54	3.96

(Continued).

Demographic Characteristic		AMAS Full Scale		LMA Subscale		MEA Subscale	
	N	Mean	SD	Mean	SD	Mean	SD
Math Test Recency							
Today	95	21.11	7.44	9.87	4.33	11.24	4.01
Yesterday	56	22.66	7.20	11.10	4.60	11.56	3.73
Earlier This Week	56	21.45	6.60	10.02	4.11	11.43	3.46
Last Week	140	22.23	7.74	10.24	4.76	11.98	3.85
Two Weeks Ago	23	21.39	9.27	10.91	5.54	10.48	4.34
More Than Two Weeks Ago	96	20.61	7.54	10.29	4.88	10.32	3.49

Table B.4

Fit Indices by Model

Model	X^2_{SB}	df	CFI[a]	TLI[a]	RMSEA[a]	CI Low	CI Up	SRMR
One-Factor Model	253.872	27	0.851	0.802	0.135	0.12	0.151	0.067
Two-Factor Model	81.346	26	0.963	0.949	0.069	0.052	0.086	0.044
Bifactor Model	28.657	18	0.993	0.985	0.037	0.000	0.061	0.019

Note. [a]Index values are corrected for robust standard errors; X^2_{SB} is chi-square test value with Satorra-Bentler (2001) correction applied.

Table B.5

Differences in Model Fit

Model	df	AIC	BIC	X^2_{SB}	ΔX^2_{SB}	Δdf	*p*-value
One-Factor Model	27	15877.18	15996.08	253.872			
Two-Factor Model	26	15655.51	15778.81	81.346	442.357	1	<.001
Bifactor Model	18	15601.46	15759.99	28.657	54.509	8	<.001

Note. X^2_{SB} is chi-square test value with Satorra-Bentler (2001) correction applied.

Table B.6

AMAS Item Loadings by Model

Indicator	One-Factor Model	Two-Factor Model		Bifactor Model		
		LMA	MEA	LMA	MEA	g
Item 1	0.507	0.489		0.110		0.504
Item 2	0.677		0.754		0.312	0.675
Item 3	0.719	0.756		0.379		0.643
Item 4	0.674		0.772		0.385	0.675
Item 5	0.478		0.576		0.464	0.441
Item 6	0.682	0.772		0.642		0.539
Item 7	0.668	0.752		0.547		0.542
Item 8	0.650		0.709		0.264	0.654
Item 9	0.667	0.663		0.235		0.630

Note. All Betas presented are standardized.

Table B.7

AMAS Item Means

Item	Mean
Item 1	1.75
Item 2	2.77
Item 3	2.32
Item 4	2.99
Item 5	3.11
Item 6	2.11
Item 7	2.09
Item 8	2.70
Item 9	2.21

Table B.8

AMAS Inter-Item Pearson Correlation Coefficients

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
Item 1								
Item 2	.357							
Item 3	.343[a]	.447						
Item 4	.313	.579[b]	.475					
Item 5	.264	.440[b]	.234	.474[b]				
Item 6	.322[a]	.393	.595[a]	.343	.199			
Item 7	.373[a]	.358	.553[a]	.330	.211	.642[a]		
Item 8	.328	.520[b]	.390	.541[b]	.417[b]	.352	.377	
Item 9	.361[a]	.398	.486[a]	.418	.309	.492[a]	.471	.440

Note. All correlation coefficients are significant at the .001 level. [a]Correlations between items loading onto the LMA scale. [b]Correlations between items loading onto the MEA scale.

Table B.9

Bifactor Model Fit Indices for Boys and Girls

Model	X^2_{SB}	df	CFI[a]	TLI[a]	RMSEA[a]	CI Low	CI Up	SRMR
Bifactor Boys and Girls Combined	31.931	18	0.989	0.979	0.045	0.017	0.070	0.022
Bifactor Model Boys	28.287	18	0.980	0.960	0.062	0.000	0.104	0.051
Bifactor Model Girls	35.864	18	0.979	0.959	0.063	0.032	0.094	0.027

Note. [a]Index values are corrected for robust standard errors; X^2_{SB} is chi-square test value with Satorra-Bentler (2001) correction applied.

Table B.10

Fit Indices of Nested Bifactor Gender Measurement Invariance Models

Invariance Step	X^2_{SB}	df	CFI[a]	TLI[a]	RMSEA[a]	RMSEA CI[a]	SRMR
Configural Invariance	64.052	36	0.980	0.959	0.063	[0.037,0.088]	0.036
Metric Invariance	82.877	51	0.977	0.968	0.056	[0.032,0.077]	0.042
Scalar Invariance	85.991	57	0.980	0.975	0.050	[0.026, 0.070]	0.042
Strict Invariance	92.329	66	0.982	0.980	0.044	[0.019, 0.064]	0.042

Note. [a]Index values are corrected for robust standard errors; X^2_{SB} is chi-square test value with Satorra-Bentler (2001) correction applied.

Table B.11

Reliability Coefficients

	AMAS Total Scale	LMA	MEA	g
α	.86	.82	.80	
ω		.83	.81	.89
ω_H				.74
ω_{HS}		.26	.20	

Table B.12

Pearson Product-Moment Correlations Between Questionnaires and AMAS Scales

Instrument	AMAS Total Scale	LMA Subscale	MEA Subscale
MARS-R	.70**	.59**	.68**
CTAS	.65**	.55**	.62**
PSWQ-C	.47**	.36**	.49**
ATMI	-.46**	-.37**	-.45**
STEM-CIS Science Scale	-.01[a]	-.06[a]	.05[a]
STEM-CIS Technology Scale	-.04[a]	-.07[a]	-.01[a]
STEM-CIS Engineering Scale	-.05[a]	-.080[a]	-.01[a]
STEM-CIS Math Scale	-.21**	-.18**	-.20**
STEM-CIS Total Scale	-.10[a]	-.12*	-.04[a]
PANAS-C-SV Positive Affect	-.17**	-.13**	-.17**

Note. $N = 422$. ** $p < .001$, * $p < .05$; [a]Correlation coefficient is not significant at $p \geq .05$.

Table B.13

Pearson Product-Moment Correlations Among Validity Questionnaires

Scale	PANAS-C-SV Positive Affect	PSWQ-C	MARS-R	ATMI	CTAS	STEM-CIS Total
PANAS-C-SV Positive Affect	1	-.302***	-.254***	.336***	-.299***	.220***
PSWQ-C	-.302***	1	.583***	-.349***	.551***	-.062
MARS-R	-.254***	.583***	1	-.561***	.635***	-.127**
ATMI	.336***	-.349***	-.561***	1	-.488***	.457***
CTAS	-.299***	.551***	.635***	-.488***	1	-.102*
STEM-CIS Total	.220***	-.062	-.127**	.457***	-.102*	1

Note. N = 422. ***p < .001, **p < .01, *p < .01.

APPENDIX C

Measures

Abbreviated Math Anxiety Scale (AMAS)

Answer each item below to indicate how much anxiety each situation causes you using the following code:

1 = low anxiety
2 = a little anxiety
3 = moderate anxiety
4 = much anxiety
5 = high anxiety

1.	Having to use tables in the back of a math book.	___	___	___	___	___
2.	Thinking about an upcoming math test 1 day before.	___	___	___	___	___
3.	Watching a teacher work on an algebraic equation on the blackboard.	___	___	___	___	___
4.	Taking an examination in a math course.	___	___	___	___	___
5.	Being given a homework assignment of many difficult problems that is due the next class meeting.	___	___	___	___	___
6.	Listening to a lecture in a math class.	___	___	___	___	___
7.	Listening to another student explain a math formula.	___	___	___	___	___
8.	Being given a "pop" quiz in a math class.	___	___	___	___	___
9.	Starting a new chapter in a math book.	___	___	___	___	___

Math Anxiety Rating Scale, Revised (MARS-R)

ATTITUDE INVENTORY

DIRECTIONS. The items in this questionnaire refer to things and experiences that may cause fear or apprehension. Answer each item below to indicate how you feel today using the following code:

1 = not at all
2 = a little
3 = moderate
4 = much
5 = very much

Work quickly and be sure to consider each item individually.

		1	2	3	4	5
1.	Watching a teacher work on an algebraic equation on the blackboard.	___	___	___	___	___
2.	Buying a textbook.	___	___	___	___	___
3.	Being given a homework assignment of many difficult problems which is due the next class meeting.	___	___	___	___	___
4.	Thinking about an upcoming math test on the day before.	___	___	___	___	___
5.	Solving a square root problem.	___	___	___	___	___
6.	Reading and interpreting graphs and charts.	___	___	___	___	___
7.	Signing up for a course in statistics.	___	___	___	___	___
8.	Listening to another student explain a math formula.	___	___	___	___	___
9.	Walking into a math class.	___	___	___	___	___
10.	Looking through the pages of a math book.	___	___	___	___	___
11.	Starting a new chapter in a math book.	___	___	___	___	___

12.	Walking on campus and thinking about a math course.	___	___	___	___	___
13.	Picking up a math textbook to begin working on a homework assignment.	___	___	___	___	___
14.	Taking an examination (quiz) in a math course.	___	___	___	___	___
15.	Reading the word "statistics".	___	___	___	___	___
16.	Working on an abstract mathematical problem such as: "If X = outstanding bills and Y = total income, calculate how much you have left for recreational expenditures".	___	___	___	___	___
17.	Reading a formula in chemistry.	___	___	___	___	___
18.	Taking an examination (final) in a math class.	___	___	___	___	___
19.	Getting ready to study for a math test.	___	___	___	___	___
20.	Being given a "pop" quiz in a math class.	___	___	___	___	___
21.	Waiting to get a math test returned on which you expected to do well.	___	___	___	___	___
22.	Listening to a lecture in a math class.	___	___	___	___	___
23.	Having to use tables in the back of a math book.	___	___	___	___	___
24.	Being told how to interpret probability statements.	___	___	___	___	___

Children's Test Anxiety Scale (CTAS)

While I am Taking Tests…	Almost Never	Some of the Time	Most of the Time	Almost Always
1. I wonder if I will pass.	1	2	3	4
2. My heart beats fast.	1	2	3	4
3. I look around the room.	1	2	3	4
4. I feel nervous.	1	2	3	4
5. I think I am going to get a bad grade.	1	2	3	4
6. It is hard for me to remember the answers.	1	2	3	4
7. I play with my pencil.	1	2	3	4
8. My face feels hot.	1	2	3	4
9. I worry about failing.	1	2	3	4
10. My belly feels funny.	1	2	3	4
11. I worry about doing something wrong.	1	2	3	4
12. I check the time.	1	2	3	4
13. I think about what my grade will be.	1	2	3	4
14. I find it hard to sit still.	1	2	3	4
15. I wonder if my answers are right.	1	2	3	4
16. I think that I should have studied more.	1	2	3	4
17. My head hurts.	1	2	3	4
18. I look at other people.	1	2	3	4

19. I think most of my answers are wrong.	1	2	3	4
20. I feel warm.	1	2	3	4
21. I worry about how hard the test is.	1	2	3	4
22. I try to finish up fast.	1	2	3	4
23. My hand shakes.	1	2	3	4
24. I think about what will happen if I fail.	1	2	3	4
25. I have to go to the bathroom.	1	2	3	4
26. I tap my feet.	1	2	3	4
27. I think about how poorly I am doing.	1	2	3	4
28. I feel scared.	1	2	3	4
29. I worry about what my parents will say.	1	2	3	4
30. I stare.	1	2	3	4

Attitudes Toward Mathematics Inventory (ATMI)

Directions: This inventory consists of statements about your attitude toward mathematics. There are no correct or incorrect responses. Read each item carefully. Please think about how you feel about each item. Enter the letter that most closely corresponds to how each statement best describes your feelings.

PLEASE USE THESE RESPONSE CODES:

A – Strongly Disagree
B – Disagree
C – Neutral
D – Agree
E – Strongly Agree

1. Mathematics is a very worthwhile and necessary subject.
2. I want to develop my mathematical skills.
3. I get a great deal of satisfaction out of solving a mathematics problem.
4. Mathematics helps develop the mind and teaches a person to think.
5. Mathematics is important in everyday life.
6. Mathematics is one of the most important subjects for people to study.
7. High school math courses would be very helpful no matter what I decide to study.
8. I can think of many ways that I use math outside of school.
9. Mathematics is one of my most dreaded subjects.
10. My mind goes blank and I am unable to think clearly when working with mathematics.
11. Studying mathematics makes me feel nervous.
12. Mathematics makes me feel uncomfortable.
13. I am always under a terrible strain in a math class.
14. When I hear the word mathematics, I have a feeling of dislike.
15. It makes me nervous to even think about having to do a mathematics problem.
16. Mathematics does not scare me at all.

17. I have a lot of self-confidence when it comes to mathematics.
18. I am able to solve mathematics problems without too much difficulty.
19. I expect to do fairly well in any math class I take.
20. I am always confused in my mathematics class.
21. I feel a sense of insecurity when attempting mathematics.
22. I learn mathematics easily.
23. I am confident that I could learn advanced mathematics.
24. I have usually enjoyed studying mathematics in school.
25. Mathematics is dull and boring.
26. I like to solve new problems in mathematics.
27. I would prefer to do an assignment in math than to write an essay.
28. I would like to avoid using mathematics in college.
29. I really like mathematics.
30. I am happier in a math class than in any other class.
31. Mathematics is a very interesting subject.
32. I am willing to take more than the required amount of mathematics.
33. I plan to take as much mathematics as I can during my education.
34. The challenge of math appeals to me.
35. I think studying advanced mathematics is useful.
36. I believe studying math helps me with problem solving in other areas.
37. I am comfortable expressing my own ideas on how to look for solutions to a difficult problem in math.
38. I am comfortable answering questions in math class.
39. A strong math background could help me in my professional life.
40. I believe I am good at solving math problems.

Penn State Worry Questionnaire for Children (PSWQ-C)

<u>Directions.</u> This form is about worrying. Worrying happens when you are scared about something and you think about it a lot. People sometimes worry about school, their family, their health, things coming up in the future, and other kinds of things. For each sentence that you read, circle the answer that best tells how true that sentence is about you.

1. My worries really bother me.	Never True	Sometimes True	Most Times True	Always True
2. I don't really worry about things.	Never True	Sometimes True	Most Times True	Always True
3. Many things make me worry.	Never True	Sometimes True	Most Times True	Always True
4. I know I shouldn't worry about things, but I just can't help it.	Never True	Sometimes True	Most Times True	Always True
5. When I am under pressure, I worry a lot.	Never True	Sometimes True	Most Times True	Always True
6. I am always worrying about something.	Never True	Sometimes True	Most Times True	Always True
7. I find it easy to stop worrying when I want.	Never True	Sometimes True	Most Times True	Always True
8. When I finish one thing, I start to worry about everything else.	Never True	Sometimes True	Most Times True	Always True
9. I never worry about anything.	Never True	Sometimes True	Most Times True	Always True
10. I've been a worrier all my life.	Never True	Sometimes True	Most Times True	Always True
11. I notice that I have been worrying about things.	Never True	Sometimes True	Most Times True	Always True

12. Once I start worrying, I can't stop.	Never True	Sometimes True	Most Times True	Always True
13. I worry all the time.	Never True	Sometimes True	Most Times True	Always True
14. I worry about things until they are all done.	Never True	Sometimes True	Most Times True	Always True

STEM Career Interest Survey (STEM-CIS)

Science

1. I am able to get a good grade in my science class.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
2. I am able to complete my science homework.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
3. I plan to use science in my future career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
4. I will work hard in my science classes.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
5. If I do well in science classes, it will help me in my future career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
6. My parents would like it if I choose a science career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
7. I am interested in careers that use science.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
8. I like my science class.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
9. I have a role model in a science career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
10. I would feel comfortable talking to people who work in science careers.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
11. I know of someone in my family who uses science in their career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

Math

12. I am able to get a good grade in my math class.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
13. I am able to complete my math homework.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
14. I plan to use mathematics in my future career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
15. I will work hard in my mathematics classes.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
16. If I do well in mathematics classes, it will help me in my future career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
17. My parents would like it if I choose a mathematics career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
18. I am interested in careers that use mathematics.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
19. I like my mathematics class.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
20. I have a role model in a mathematics career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
21. I would feel comfortable talking to people who work in mathematics careers.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
22. I know of someone in my family who uses mathematics in their career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

Technology

23. I am able to do well in activities that involve technology.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
24. I am able to learn new technologies.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
25. I plan to use technology in my future career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
26. I will learn about new technologies that will help me with school.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
27. If I learn a lot about technology, I will be able to do lots of different types of careers.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
28. My parents would like it if I choose a technology career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
29. I like to use technology for class work.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
30. I am interested in careers that use technology.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
31. I have a role model who uses technology in their career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
32. I would feel comfortable talking to people who work in technology careers.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
33. I know of someone in my family who uses technology in their career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

Engineering

34. I am able to do well in activities that involve engineering.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
35. I am able to complete activities that involve engineering.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
36. I plan to use engineering in my future career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
37. I will work hard on activities at school that involve engineering.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
38. If I learn a lot about engineering, I will be able to do lots of different types of careers.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
39. My parents would like it if I choose an engineering career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
40. I am interested in careers that involve engineering.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
41. I like activities that involve engineering.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
42. I have a role model in an engineering career.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
43. I would feel comfortable talking to people who are engineers.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
44. I know of someone in my family who is an engineer.	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

Positive and Negative Affect Scale for Children, Revised Version (PANAS-C-RV)

This scale consists of a number of words that describe different feelings and emotions. Read each item and then circle the appropriate answer next to that word.

Indicate to what extent you have felt this way during the past few weeks.

Feeling or emotion	Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
Joyful	1	2	3	4	5
Cheerful	1	2	3	4	5
Happy	1	2	3	4	5
Lively	1	2	3	4	5
Proud	1	2	3	4	5
Miserable	1	2	3	4	5
Mad	1	2	3	4	5
Afraid	1	2	3	4	5
Scared	1	2	3	4	5
Sad	1	2	3	4	5

Demographic Questionnaire

Demographic Information

Do not write your name on this page or any page in this packet.
Please answer the questions in this packet to the best of your ability. Try to answer all questions. If you do not know the answer, try to guess.

What is your age? _________________

What grade are you in? _________________

What's the name of your school? _________________

What is your gender? Boy Girl

What is your race/ethnicity? (Circle all that apply)

White Black Hispanic Asian Other_________________

What is the highest education attained by your mother or father?

Some high school or less

High school diploma or GED

Vocational school or some college

College degree

Professional or graduate degree (for example, doctor, lawyer, professor)

What was your grade last semester in the following school subjects? (Please circle)

History/Social Studies	A	B	C	D	F
Math	A	B	C	D	F
English/Reading	A	B	C	D	F
Science	A	B	C	D	F

Have you ever been diagnosed with by any of the following by a health provider? (for example, a doctor or psychologist)

ADHD or ADD

Autism

Anxiety

Depression

Epilepsy/Seizure Disorder

Learning Disability/Disorder

If yes, do you know what kind? _________________

Have you ever been enrolled in special education classes at school?

No

Yes

During which grades? __________________________________

For which classes/subjects? __________________________________

Were you taught in a separate classroom or with the rest of your grade?

____

When was the last time you took a <u>math</u> quiz or test? (This can be for class or a standardized test.)

Today

Yesterday

Earlier this week

Last week

Two weeks ago

More than two weeks ago

REFERENCES

Ahmed, W. (2018). Developmental trajectories of math anxiety during adolescence: Associations with STEM career choice. *Journal of Adolescence*, *67*, 158–166. https://doi.org/10.1016/j.adolescence.2018.06.010

Alexander, L., & Cobb, R. (1987). Identification of the dimensions and predictors of math anxiety among college students. *Journal of Human Behavior & Learning*, *4*(1), 25–32.

Alexander, L., & Martray, C. (1989). The Development of an Abbreviated Version of the Mathematics Anxiety Rating Scale. *Measurement and Evaluation in Counseling and Development*, *22*(3), 143–150. https://doi.org/10.1080/07481756.1989.12022923

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype Susceptibility in Children: Effects of Identity Activation on Quantitative Performance. *Psychological Science*, *12*(5), 385–390. https://doi.org/10.1111/1467-9280.00371

Anderson, M., & Jiang, J. (2018). *Teens, Social Media & Technology 2018: Methodology*. Pew Research Center. https://www.pewresearch.org/internet/2018/05/31/teens-technology-methodology/

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

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

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*(2), 224–237. https://doi.org/10.1037//0096-3445.130.2.224

Ashcraft, M. H., & Ridley, K. S. (2005). Math Anxiety and Its Cognitive Consequences: A Tutorial Review. In J. I. D. Campbell, *Handbook of Mathematical Cognition*. Psychology Press.

Baloğlu, M., & Zelhart, P. F. (2007). Psychometric Properties of the Revised Mathematics Anxiety Rating Scale. *The Psychological Record*, *57*(4), 593–611. https://doi.org/10.1007/BF03395597

Bandalos, D. L. (2018). *Measurement Theory and Applications for the Social Sciences*. Guilford Publications.

Bardhoshi, G., & Erford, B. T. (2017). Processes and Procedures for Estimating Score Reliability and Precision. *Measurement and Evaluation in Counseling and Development*, *50*(4), 256–263. https://doi.org/10.1080/07481756.2017.1388680

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246.

Bentler, P. M. (1995). *EQS Structural Equations Program Manual*. Multivariate Software.

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*. https://doi.org/10.3389/fpsyg.2015.01404

Blackburn, H. (2017). The Status of Women in STEM in Higher Education: A Review of the Literature 2007–2017. *Science & Technology Libraries*, *36*(3), 235–273. https://doi.org/10.1080/0194262X.2017.1371658

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate Use of Bifactor Analysis in Psychopathology Research: Appreciating Benefits and Limitations. *Biological Psychiatry*, *88*(1), 18–27. https://doi.org/10.1016/j.biopsych.2020.01.013

Brown, T. A. (2014). *Confirmatory Factor Analysis for Applied Research* (Second). Guilford Publications.

Brunyé, T. T., Mahoney, C. R., Giles, G. E., Rapp, D. N., Taylor, H. A., & Kanarek, R. B. (2013). Learning to relax: Evaluating four brief interventions for overcoming the negative emotions accompanying math anxiety. *Learning and Individual Differences*, *27*, 1–7. https://doi.org/10.1016/j.lindif.2013.06.008

Brush, L. R. (1978). A Validation Study of the Mathematics Anxiety Rating Scale (Mars. *Educational and Psychological Measurement*, *38*(2), 485–499. https://doi.org/10.1177/001316447803800236

Buffington, C., Harris, B. C., Jones, C., & Weinberg, B. A. (2016). STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data linked to the 2010 Census. *The American Economic Review*, *106*(5), 333–338. https://doi.org/10.1257/aer.p20161124

Canu, W. H., Elizondo, M., & Broman-Fulks, J. J. (2017). History of ADHD traits related to general test and specific math anxiety in college students. *Learning and Individual Differences*, *58*, 56–63. https://doi.org/10.1016/j.lindif.2017.07.008

Capraro, M. M., Capraro, R. M., & Henson, R. K. (2001). Measurement Error of Scores on the Mathematics Anxiety Rating Scale across Studies. *Educational and Psychological Measurement*, *61*(3), 373–386. https://doi.org/10.1177/00131640121971266

Carey, E., Hill, F., Devine, A., & Szücs, D. (2016). The Chicken or the Egg? The Direction of the Relationship Between Mathematics Anxiety and Mathematics Performance. *Frontiers in Psychology*, *6*. https://doi.org/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*. https://doi.org/10.3389/fpsyg.2017.00011

Casanova, S., Vukovic, R. K., & Kieffer, M. J. (2021). Do girls pay an unequal price? Black and Latina girls' math attitudes, math anxiety, and mathematics achievement. *Journal of Applied Developmental Psychology*, *73*, 101256. https://doi.org/10.1016/j.appdev.2021.101256

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

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling General and Specific Variance in Multifaceted Constructs: A Comparison of the Bifactor Model to Other Approaches. *Journal of Personality*, *80*(1), 219–251. https://doi.org/10.1111/j.1467-6494.2011.00739.x

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A Comparison of Bifactor and Second-Order Models of Quality of Life. *Multivariate Behavioral Research*, *41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Chen, T., Xu, M., Tu, J., Wang, H., & Niu, X. (2018). Relationship between Omnibus and Post-hoc Tests: An Investigation of performance of the F test in ANOVA. *Shanghai Archives of Psychiatry*, *30*(1), 60–64. https://doi.org/10.11919/j.issn.1002-0829.218014

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chorpita, B. F., Tracey, S. A., Brown, T. A., Collica, T. J., & Barlow, D. H. (1997). Assessment of worry in children and adolescents: An adaptation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, *35*(6), 569–581.

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*. https://doi.org/10.3389/fpsyg.2015.01833

Cisler, J. M., & Koster, E. H. W. (2010). Mechanisms of Attentional Biases towards Threat in the Anxiety Disorders: An Integrative Review. *Clinical Psychology Review*, *30*(2), 203. https://doi.org/10.1016/j.cpr.2009.11.003

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Conlon, R. A., Hicks, A., Barroso, C., & Ganley, C. M. (2021). The effect of the timing of math anxiety measurement on math outcomes. *Learning and Individual Differences*, *86*, 101962. https://doi.org/10.1016/j.lindif.2020.101962

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

D'Ailly, H., & Bergering, A. J. (1992). Mathematics Anxiety and Mathematics Avoidance Behavior: A Validation Study of Two MARS Factor-Derived Scales. *Educational and Psychological Measurement*, *52*(2), 369–377. https://doi.org/10.1177/0013164492052002012

DeMars, C. E. (2013). A Tutorial on Interpreting Bifactor Model Scores. *International Journal of Testing*, *13*(4), 354–378. https://doi.org/10.1080/15305058.2013.799067

Dempster, A. P., Laird, N. M., & work(s):, D. B. R. R. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.

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*, 33. https://doi.org/10.1186/1744-9081-8-33

Dew, K. M. H., Galassi, J. P., & Galassi, M. D. (1983). Mathematics Anxiety: Some Basic Issues. *Journal of Counseling Psychology*, *30*(3), 443–446. http://dx.doi.org/10.1037/0022-0167.30.3.443

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

Dreger, R. M., & Aiken, L. R., Jr. (1957). The identification of number anxiety in a college population. *Journal of Educational Psychology*, *48*(6), 344–351. https://doi.org/10.1037/h0045894

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

Ebesutani, C., Regan, J., Smith, A., Reise, S., Higa-McMillan, C., & Chorpita, B. F. (2012). The 10-Item Positive and Negative Affect Schedule for Children, Child and Parent Shortened Versions: Application of Item Response Theory for More Efficient Assessment. *Journal of Psychopathology and Behavioral Assessment*, *34*(2), 191–203. https://doi.org/10.1007/s10862-011-9273-2

Fayer, S., Lacey, A., & Watson, A. (2017). *STEM Occupations: Past, Present, And Future* (Spotlight on Statistics, p. 35). U.S. Bureau of Labor Statistics. https://www.bls.gov/spotlight/2017/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future/pdf/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future.pdf

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman Mathematics Attitudes Scales: Instruments Designed to Measure Attitudes toward the Learning of Mathematics by Females and Males. *Journal for Research in Mathematics Education*, *7*(5), 324. https://doi.org/10.2307/748467

Fernández, L. M., Wang, X., Ramirez, O., & Villalobos, M. C. (2021). Latinx Students' Mathematics Anxiety and Their Study Habits: Exploring Their Relationship at the Postsecondary Level. *Journal of Hispanic Higher Education*, *20*(3), 278–296. https://doi.org/10.1177/1538192719852205

Funk, C., & Parker, K. (2018). *Women and Men in STEM Often at Odds Over Workplace Equity*. Pew Research Center. https://www.pewresearch.org/social-trends/2018/01/09/blacks-in-stem-jobs-are-especially-concerned-about-diversity-and-discrimination-in-the-workplace/

Furtner, M. R., Rauthmann, J. F., & Sachse, P. (2015). Unique self-leadership: A bifactor model approach. *Leadership*, *11*(1), 105–125. https://doi.org/10.1177/1742715013511484

Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, *73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Georgiou, A., Soulis, S.-G., Rapti, D., & Papanikolaou, K. (2018). Math Anxiety of Students with High Functioning Autism Spectrum Disorder. *American International Journal of Social Science*, *7*(4). https://doi.org/10.30845/aijss.v7n4p15

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do Girls Really Experience More Anxiety in Mathematics? *Psychological Science*, *24*(10), 2079–2087. https://doi.org/10.1177/0956797613486989

Golberstein, E., Wen, H., & Miller, B. F. (2020). Coronavirus Disease 2019 (COVID-19) and Mental Health for Children and Adolescents. *JAMA Pediatrics*, *174*(9), 819. https://doi.org/10.1001/jamapediatrics.2020.1456

Gough, S. M. F. (1954). Why Failures in Mathematics? Mathemaphobia: Causes and Treatments. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *28*(5), 290–294. https://doi.org/10.1080/00098655.1954.11476830

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. https://doi.org/10.1111/j.1524-4733.2007.00231.x

Hammer, J. H., & Toland, M. D. (2016, November). *Bifactor analysis in Mplus* [Video file]. http://sites.education.uky.edu/apslab/upcoming-events/

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

Hart, S. A., Logan, J. A. R., Thompson, L., Kovas, Y., McLoughlin, G., & Petrill, S. A. (2016). A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *Journal of Educational Psychology*, *108*(2), 181–193. https://doi.org/10.1037/edu0000045

Hatcher, L. (1994). *A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling*. SAS Institute, Inc., Cary.

Hembree, R. (1990). The Nature, Effects, and Relief of Mathematics Anxiety. *Journal for Research in Mathematics Education*, *21*(1), 33. https://doi.org/10.2307/749455

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

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, *6*(1), 8. https://doi.org/10.21427/D7CF7R

Hopko, D. R. (2003). Confirmatory Factor Analysis Of The Math Anxiety Rating Scale–Revised. *Educational and Psychological Measurement*, *63*(2), 336–351. https://doi.org/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. https://doi.org/10.1177/1073191103010002008

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

Huang, X., Zhang, J., & Hudson, L. (2019). Impact of math self-efficacy, math anxiety, and growth mindset on math and science career interest for middle school students: The gender moderating effect. *European Journal of Psychology of Education*, *34*(3), 621–640. https://doi.org/10.1007/s10212-018-0403-z

Hughes, A. A., & Kendall, P. C. (2009). Psychometric Properties of the Positive and Negative Affect Scale for Children (PANAS-C) in Children with Anxiety Disorders. *Child Psychiatry and Human Development*, *40*(3), 343–352. https://doi.org/10.1007/s10578-009-0130-4

Jamieson, J. P., Peters, B. J., Greenwood, E. J., & Altose, A. J. (2016). Reappraising Stress Arousal Improves Performance and Reduces Evaluation Anxiety in Classroom Exam Situations. *Social Psychological and Personality Science*, *7*(6), 579–587. https://doi.org/10.1177/1948550616644656

Jones, E. A. K., Mitra, A. K., & Bhuiyan, A. R. (2021). Impact of COVID-19 on Mental Health in Adolescents: A Systematic Review. *International Journal of Environmental Research and Public Health*, *18*(5). https://doi.org/10.3390/ijerph18052470

Kazelskis, R., Reeves, C., Kersh, M. E., Bailey, G., Cole, K., Larmon, M., Hall, L., & Holliday, D. C. (2000). Mathematics Anxiety and Test Anxiety: Separate Constructs? *The Journal of Experimental Education*, *68*(2), 137–146. https://doi.org/10.1080/00220970009598499

Keene, O. N. (1995). The log transformation is special. *Statistics in Medicine*, *14*(8), 811–819. https://doi.org/10.1002/sim.4780140810

Kier, M. W., Blanchard, M. R., Osborne, J. W., & Albert, J. L. (2014). The Development of the STEM Career Interest Survey (STEM-CIS). *Research in Science Education*, *44*(3), 461–481. https://doi.org/10.1007/s11165-013-9389-3

Kim, H.-Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*(1), 52–54. https://doi.org/10.5395/rde.2013.38.1.52

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the Potential Impact of COVID-19 School Closures on Academic Achievement. *Educational Researcher*, *49*(8), 549–565. https://doi.org/10.3102/0013189X20965918

Larson, N., Neumark-Sztainer, D., Harwood, E. M., Eisenberg, M. E., Wall, M. M., & Hannan, P. J. (2011). Do young adults participate in surveys that 'go green'? Response rates to a web and mailed survey of weight-related health behaviors. *International Journal of Child Health and Human Development : IJCHD*, *4*(2), 225–231.

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*(1), 79–122. https://doi.org/10.1006/jvbe.1994.1027

Levene, H. (1960). Robust Tests for Equality of Variances. In I. Olkin (Ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (pp. 278–292). Stanford University Press.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, *136*(6), 1123–1135. https://doi.org/10.1037/a0021276

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Luo, X., Wang, F., & Luo, Z. (2009). Investigation and Analysis of Mathematics Anxiety in Middle School Students. *Journal of Mathematics Education*, *2*(2), 12–19.

Luttenberger, S., Wimmer, S., & Paechter, M. (2018). Spotlight on math anxiety. *Psychology Research and Behavior Management*, *Volume 11*, 311–322. https://doi.org/10.2147/PRBM.S141421

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. https://doi.org/10.2307/749772

Ma, X., & Xu, J. (2004). Determining the Causal Ordering between Attitude toward Mathematics and Achievement in Mathematics. *American Journal of Education*, *110*(3), 256–280. https://doi.org/10.1086/383074

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences*, *16*(8), 404–406. https://doi.org/10.1016/j.tics.2012.06.008

Maloney, E. A., Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2015). Intergenerational Effects of Parents' Math Anxiety on Children's Math Achievement and Anxiety. *Psychological Science*, *26*(9), 1480–1488. https://doi.org/10.1177/0956797615592630

Mangunkusumo, R. T., Moorman, P. W., Van Den Berg-de Ruiter, A. E., Van Der Lei, J., De Koning, H. J., & Raat, H. (2005). Internet-administered adolescent health questionnaires compared with a paper version in a randomized study. *The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine*, *36*(1), 70.e1-6. https://doi.org/10.1016/j.jadohealth.2004.02.020

Markon, K. E. (2019). Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annual Review of Clinical Psychology*, *15*(1), 51–69. https://doi.org/10.1146/annurev-clinpsy-050718-095522

Marques de Miranda, D., da Silva Athanasio, B., Sena Oliveira, A. C., & Simoes-e-Silva, A. C. (2020). How is COVID-19 pandemic impacting mental health of children and adolescents? *International Journal of Disaster Risk Reduction*, *51*, 101845. https://doi.org/10.1016/j.ijdrr.2020.101845

McAuliffe, E. A., & Trueblood, C. (1986, April). *Factor Analysis: A Tool For Studying Mathematics Anxiety* [Paper]. American Educational Research Association, San Francisco, CA.

McDonald, R. P. (1999). *Test Theory: A Unified Treatment* (1st ed.). Psychology Press.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology*, *82*(1), 60–70. https://doi.org/10.1037/0022-0663.82.1.60

Meredith, W., & Teresi, J. A. (2006). An Essay on Measurement and Factorial Invariance. *Medical Care*, *44*(Suppl 3), S69–S77. https://doi.org/10.1097/01.mlr.0000245438.73837.89

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive Statistics and Normality Tests for Statistical Data. *Annals of Cardiac Anaesthesia*, *22*(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18

Muehlhausen, W., Doll, H., Quadri, N., Fordham, B., O'Donohoe, P., Dogar, N., & Wild, D. J. (2015). Equivalence of electronic and paper administration of patient-reported outcome measures: A systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health and Quality of Life Outcomes*, *13*(1), 167. https://doi.org/10.1186/s12955-015-0362-x

National Center for Science and Engineering Statistics. (2021). *Women, Minorities, and Persons with Disabilities in Science and Engineering*. National Science Foundation. https://ncses.nsf.gov/pubs/nsf21321/report

National Science Board. (2018). *Chapter 3: Science and Engineering Labor Force* (Science & Engineering Indicators 2018). https://www.nsf.gov/statistics/2018/nsb20181/assets/901/science-and-engineering-labor-force.pdf

Núñez-Peña, M. I., Suárez-Pellicioni, M., & Bono, R. (2013). Effects of math anxiety on student success in higher education. *International Journal of Educational Research*, *58*, 36–43. https://doi.org/10.1016/j.ijer.2012.12.004

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

O'Leary, K., Fitzpatrick, C. L., & Hallett, D. (2017). Math Anxiety Is Related to Some, but Not All, Experiences with Math. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02067

Ortiz, R., Kishton, R., Sinko, L., Fingerman, M., Moreland, D., Wood, J., & Venkataramani, A. (2021). Assessing Child Abuse Hotline Inquiries in the Wake of COVID-19: Answering the Call. *JAMA Pediatrics*. https://doi.org/10.1001/jamapediatrics.2021.0525

Osborne, J. (2001). Testing Stereotype Threat: Does Anxiety Explain Race and Sex Differences in Achievement? *Contemporary Educational Psychology*, *26*, 291–310. https://doi.org/10.1006/ceps.2000.1052

Park, D., Ramirez, G., & Beilock, S. L. (2014). The role of expressive writing in math anxiety. *Journal of Experimental Psychology. Applied*, *20*(2), 103–111. https://doi.org/10.1037/xap0000013

Passolunghi, M. C., Caviola, S., De Agostini, R., Perin, C., & Mammarella, I. C. (2016). Mathematics Anxiety, Working Memory, and Mathematics Performance in Secondary-School Children. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00042

Pestle, S. L., Chorpita, B. F., & Schiffman, J. (2008). Psychometric Properties of the Penn State Worry Questionnaire for Children in a Large Clinical Sample. *Journal of Clinical Child & Adolescent Psychology*, *37*(2), 465–471. https://doi.org/10.1080/15374410801955896

Piedmont, R. L. (2014). Inter-item Correlations. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3303–3304). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1493

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*(2), 551–557. https://doi.org/10.1177/001316448204200218

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

Putnick, D. L., & Bornstein, M. H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Developmental Review : DR*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology*, *141*, 83–100. https://doi.org/10.1016/j.jecp.2015.07.014

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math Anxiety, Working Memory, and Math Achievement in Early Elementary School. *Journal of Cognition and Development*, *14*(2), 187–202. https://doi.org/10.1080/15248372.2012.664593

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math Anxiety: Past Research, Promising Interventions, and a New Interpretation Framework. *Educational Psychologist*, *53*(3), 145–164. https://doi.org/10.1080/00461520.2018.1447384

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to Psychometric Theory*. Routledge.

Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology*, *107*(3), 645–662. https://doi.org/10.1037/edu0000012

Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Resnick, H., Viehe, J., & Segal, S. (1982). Is Math Anxiety a Local Phenomenon? A Study of Prevalence and Dimensionality. *Journal of Counseling Psychology*, *29*(1), 39–47.

Reynolds, C. R., & Richmond, B. O. (1985). *Revised Children's Manifest Anxiety Scales (RCMAS)*. Western Psychological Services.

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

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Rolstad, S., Adler, J., & Rydén, A. (2011). Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis. *Value in Health*, *14*(8), 1101–1108. https://doi.org/10.1016/j.jval.2011.06.003

Rounds, J. B., & Hendel, D. D. (1980). Measurement and dimensionality of mathematics anxiety. *Journal of Counseling Psychology*, *27*(2), 138–149. https://doi.org/10.1037/0022-0167.27.2.138

Rubinsten, O., Eidlin, H., Wohl, H., & Akibli, O. (2015). Attentional bias in math anxiety. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01539

Rubinsten, O., & Tannock, R. (2010). Mathematics anxiety in children with developmental dyscalculia. *Behavioral and Brain Functions*, *6*(46), 13. https://doi-org/10.1186/1744-9081-6-46

Sahlqvist, S., Song, Y., Bull, F., Adams, E., Preston, J., & Ogilvie, D. (2011). Effect of questionnaire length, personalisation and reminder type on response rate to a complex postal survey: Randomised controlled trial. *BMC Medical Research Methodology*, *11*, 62. https://doi.org/10.1186/1471-2288-11-62

Samuel, T., & Warner, J. (2019). "I Can Math!": Reducing Math Anxiety and Increasing Math Self-Efficacy Using a Mindfulness and Growth Mindset-Based Intervention in First-Year Students. *Community College Journal of Research and Practice*, *45*, 1–18. https://doi.org/10.1080/10668926.2019.1666063

Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Shevlin, M., & Miles, J. N. V. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences*, *25*(1), 85–90. https://doi.org/10.1016/S0191-8869(98)00055-5

Sidak, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, *62*(318), 626–633. https://doi.org/10.2307/2283989

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28. https://doi.org/10.1006/jesp.1998.1373

Spielberger, C. D., Gonzalez, H. P., Taylor, C. J., Anton, E. D., Ross, G. R., & Westberry, L. G. (1980). *Manual for the Test Anxiety Inventory ("Test Attitude Inventory")*. Consulting Psychologist Press.

Steele, C. M. (1997). A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance. *American Psychologist*, *52*(6), 17. http://dx.doi.org/10.1037/0003-066X.52.6.613

Steele, J. (2003). Children's Gender Stereotypes About Math: The Role of Stereotype Stratification. *Journal of Applied Social Psychology*, *33*(12), 2587–2606. https://doi.org/10.1111/j.1559-1816.2003.tb02782.x

Steiger, J. H., & Lind, J. M. (1980, June). *Statistically based tests for the number of common factors* [Paper]. Statistically based tests for the number of common factors, Iowa City, IA.

Suárez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, À. (2015). Attentional bias in high math-anxious individuals: Evidence from an emotional Stroop task. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01577

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. https://doi.org/10.3758/s13415-015-0370-7

Suinn, R. M. (1969). The STABS, a measure of test anxiety for behavior therapy: Normative data. *Behaviour Research and Therapy*, *7*(3), 335–339. https://doi.org/The STABS, a measure of test anxiety for behavior therapy

Supekar, K., Iuculano, T., Chen, L., & Menon, V. (2015). Remediation of Childhood Math Anxiety and Associated Neural Circuits through Cognitive Tutoring. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(36), 12574–12583. https://doi.org/10.1523/JNEUROSCI.0786-15.2015

Tapia, M. (1996, November 6). *The Attitudes toward Mathematics Instrument* [Paper]. Annual Meeting of the Mid-South Educational Research Association, Tuscaloosa, AL. http://eric.ed.gov/?id=ED404165

Tapia, M., & Marsh, G. E. (2000, November 16). *Attitudes Toward Mathematics Instrument: An Investigation with Middle School Students* [Paper]. Annual Meeting of the Mid-South Educational Research Association, Bowling Green, KY. https://files-eric-ed-gov.ezproxy.baylor.edu/fulltext/ED449045.pdf

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/BF02291170

Tukey, J. W. (1977). Exploratory data analysis. In *Addison-Wesley Seriesin Behavioral Science: Quantitative Methods*. Addison-Wesley. http://adsabs.harvard.edu/abs/1977eda..book.....T

Ünlü, Z. K., Dököe, İ., & Ünlü, V. (2016). Adaptation of the Science, Technology, Engineering, and Mathematics Career Interest Survey (STEM-CIS) into Turkish. *Eurasian Journal of Educational Research*, *16*(63), 21–36. https://doi.org/10.14689/ejer.2016.63.2

U.S. Department of Education. (2021). *Science, Technology, Engineering, and Math, including Computer Science*. https://www.ed.gov/stem

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.

Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). *Model Comparison and the Principle of Parsimony* (J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels, Eds.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199957996.013.14

Wang, Z., Hart, S. A., Kovas, Y., Lukowski, S., Soden, B., Thompson, L. A., Plomin, R., McLoughlin, G., Bartlett, C. W., Lyons, I. M., & Petrill, S. A. (2014). Who is afraid of math? Two sources of genetic variance for mathematical anxiety. *Journal of Child Psychology and Psychiatry*, *55*(9), 1056–1064. https://doi.org/10.1111/jcpp.12224

Wigfield, A., & Meece, J. L. (1988). Math Anxiety in Elementary and Secondary School Students. *Journal of Educational Psychology*, *80*(2), 210–216. http://dx.doi.org/10.1037/0022-0663.80.2.210

Wren, D. G., & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress & Coping*, *17*(3), 227–240. https://doi.org/10.1080/10615800412331292606

Wu, S. S., Willcutt, E. G., Escovar, E., & Menon, V. (2014). Mathematics Achievement and Anxiety and Their Relation to Internalizing and Externalizing Behaviors. *Journal of Learning Disabilities*, *47*(6), 503–514. https://doi.org/10.1177/0022219412473154

Yang, Y., & Green, S. B. (2011). Coefficient Alpha: A Reliability Coefficient for the 21st Century? *Journal of Psychoeducational Assessment*, *29*(4), 377–392. https://doi.org/10.1177/0734282911406668

Zwaanswijk, W., Veen, V. C., & Vedder, P. (2017). The Youth Psychopathic Traits Inventory: A Bifactor Model, Dimensionality, and Measurement Invariance. *Assessment*, *24*(7), 932–944. https://doi.org/10.1177/1073191116632340